Collection of Biostatistics Research Archive
COBRA Preprint Series

Year 2006 *Paper* 15

Properties of Monotonic Effects

Tyler J. VanderWeele* James M. Robins†

*University of Chicago, tvanderw@hsph.harvard.edu
†Harvard School of Public Health, robins@hsph.harvard.edu

Properties of Monotonic Effects

Tyler J. VanderWeele and James M. Robins

Abstract

Various relationships are shown hold between monotonic effects and weak monotonic effects and the monotonicity of certain conditional expectations. This relationship is considered for both binary and non-binary variables. Counterexamples are provide to show that the results do not hold under less restrictive conditions. The ideas of monotonic effects are furthermore used to relate signed edges on a directed acyclic graph to qualitative effect modification.

1. Introduction

The concept of monotonic effects has proven useful in determining the sign of the bias that arises when control for confounding is inadequate (VanderWeele and Robins, 2006a), in determining the sign of the covariance and conditional covariance amongst variables (VanderWeele and Robins, 2006a, 2006b) and in constructing tests for detecting the presence of synergism (VanderWeele and Robins, 2006c). The relation of monotonic effects to problems in causal inference is presented elsewhere (VanderWeele and Robins, 2006a) and two results relating monotonic effects to causal inference are reviewed in Appendix 1. In this paper, however, we will develop a number of statistical properties concerning monotonic effects and weak monotonic effects. The paper is organized as follows. In Section 2 we present the notation we will use in this paper and review the definitions concerning directed acyclic graphs. In Section 3, we introduce the concepts of a monotonic effect and a weak monotonic effect in the directed acyclic graph framework. In Section 4, we present and develop a number of results relating weak monotonic effects to the monotonicity in the conditioning argument of certain conditional expectations. Finally, in Section 5, we develop a number of results that relate weak monotonic effects to the existence of qualitative effect modifiers.

2. Notation and Directed Acyclic Graphs

Following Pearl (1995), a causal directed acyclic graph is a set of nodes $(X_1, ..., X_n)$ and directed edges amongst nodes such that the graph has no cycles and such that for each node X_i on the graph the corresponding variable is given by its non-parametric structural equation $X_i = f_i(pa_i, \epsilon_i)$ where pa_i are the parents of X_i on the graph and the ϵ_i are mutually independent. These non-parametric structural equations can be seen as a generalization of the path analysis and linear structural equation models (Pearl 1995, 2000) developed by Wright (1921) in the genetics literature and Haavelmo (1943) in the econometrics literature. Directed acyclic graphs can be interpreted as represent causal relationships. The non-parametric structural equations encode counterfactual relationships amongst the variables represented on the graph. The equations themselves represent one-step ahead counterfactuals with other counterfactuals given by recursive substitution. The requirement that the ϵ_i be mutually independent is essentially a requirement that there is no variable absent from the graph which, if included on the graph, would be a parent of two or more variables (Pearl, 1995, 2000). Further discussion of the causal interpretation of directed acyclic graphs can be found elsewhere (Pearl, 1995, 2000; Greenland et al. 1999; Dawid 2002; Spirtes 2002; Robins 2003).

A path is a sequence of nodes connected by edges regardless of arrowhead direction; a directed path is a path which follows the edges in the direction indicated by the graph's arrows. A node C is said to be a common cause of A and Y if there exists a directed path from C to Y not through A and a directed path from C to A not through Y. A collider is a particular node on a path such that both the preceding and subsequent nodes on the path have directed edges going into that node i.e. both the edge to and the edge from that node have arrowheads into the node. A path between A and B is said to be blocked given some set of variables Z if either there is a variable in Z on the path that is not a collider or if there is a collider on the path such that neither the collider itself nor any of its descendants are in Z. It has been shown that if all paths between A and B are blocked given Z then A and B are conditionally independent given Z (Verma and Pearl, 1988; Geiger et al., 1990; Lauritzen et al., 1990). The directed acyclic graph causal framework has proven to be particularly useful in determining whether conditioning on a given set of variables, or none at all, is sufficient to control for confounding. The most important result in this regard is the back-door path criterion (Pearl, 1995). A back-door path from some node A to another node Y is a path which begins with a directed edge into A. Pearl (1995) showed that for intervention variable A and outcome Y, if a set of variables Z is such that no variable in Z is a descendent of A and such that Z blocks all back-door paths from A to Y then conditioning on Z suffices to control for confounding for the estimation of the causal effect of A on Y. The counterfactual value of Y intervening to set $A = a$ we denote by $Y_{A=a}$.

3. On the Definition of a Monotonic Effect

The definition of a monotonic effect is given in terms of a directed acyclic graph's nonparametric structural equations.

Definition 1. The non-parametric structural equation for some node Y on a causal directed acyclic graph with parent A can be expressed as $Y = f(\widetilde{pa}_A, A, \epsilon_Y)$ where $\widetilde{pa}_Y$ are the parents of Y other than A; A is said to have a positive monotonic effect on Y if for all $\widetilde{pa}_Y$ and ϵ_Y, $f(\widetilde{pa}_Y, A_1, \epsilon_Y) \geq f(\widetilde{pa}_Y, A_2, \epsilon_Y)$ whenever $A_1 \geq A_2$. Similarly A is said to have a negative monotonic effect on Y if for all $\widetilde{pa}_Y$ and ϵ_Y, $f(\widetilde{pa}_Y, A_1, \epsilon_Y) \leq f(\widetilde{pa}_Y, A_2, \epsilon_Y)$ whenever $A_1 \geq A_2$.

The presence of a monotonic effects is closely related to the monotonicity of counterfactual variables as is made clear by the following proposition. All proofs of all propositions and theorems are given in Appendix 2.

Proposition 1. The variable A has a positive monotonic effect on Y if and only if for all individuals ω in the population and all values of $\widetilde{pa}_Y$, $Y_{a_1,\widetilde{pa}_Y}(\omega) \geq Y_{a_0,\widetilde{pa}_Y}(\omega)$ whenever $a_1 \geq a_0$.

Because for any individual we observe the outcome only under one particular value of the intervention variable, the presence of a monotonic effect is not identifiable. The results presented in this paper are in fact true under slightly weaker conditions which are identifiable when data on all of the directed acyclic graph's variables are observed. We thus introduce the concept of a weak monotonic effect.

Definition 2. Suppose that variable A is a parent of some variable Y and let $\widetilde{pa}_Y$ denote the parents of Y other than A. We say that A has a *weak positive monotonic effect* on Y if the survivor function $S(y|a, \widetilde{pa}_Y) = P(Y \geq y|A = a, \widetilde{pa}_Y)$ is such that whenever $a_1 \geq a_0$ we have $S(y|a_1, \widetilde{pa}_Y) \geq S(y|a_0, \widetilde{pa}_Y)$ for all y and all $\widetilde{pa}_Y$; the variable A is said to have a *weak negative monotonic effect* on Y if whenever $a_1 \geq a_0$ we have $S(y|a_1, \widetilde{pa}_Y) \leq S(y|a_0, \widetilde{pa}_Y)$ for all y and all $\widetilde{pa}_Y$.

We note that for parent A and child Y, the definition of a weak monotonic effect coincides with Wellman's (1990) definition of positive qualitative influence when the "context" for qualitative influence is chosen to be the parents of Y other than A.

Proposition 2. If A has a positive monotonic effect on Y then A has a weak positive monotonic effect on Y.

A monotonic effect is a relation between two nodes on a directed acyclic graph and as such it is associated with an edge. The definition of the sign of an edge can be given either in terms of monotonic effects or weak monotonic effects. We can define the sign of an edge as the sign of the monotonic effect or weak monotonic effect to which the edge corresponds; this in turn gives rise to a natural definition for the sign of a path.

Definition 3. An edge on a causal directed acyclic graph from X to Y is said to be of positive sign if X has a positive monotonic effect on Y. An edge from X to Y is said to be of negative sign if X has a negative monotonic effect on Y. If X has neither a positive monotonic effect nor a negative monotonic effect on Y, then the edge from X to Y is said to be without a sign.

Definition 4. The sign of a path on a causal directed acyclic graph is the product of the signs of the edges that constitute that path. If one of the edges on a path is without a sign then the sign of the path is said to be undefined.

We will call a causal directed acyclic graph with signs on those edges which allow them a signed causal directed acyclic graph. The theorems in this paper are given in terms of signed paths so as to be applicable to both monotonic effects and weak monotonic effects. One further definition will be useful in the development of the theory below.

Definition 5. Two variables X and Y are said to be positively monotonically associated if all directed paths between X and Y are of positive sign and all common causes C_i of X and Y are such that all directed paths from C_i to X are of the same sign as all directed paths from C_i to Y; the variables X and Y are said to be negatively monotonically associated if all directed paths between X and Y are of negative sign and all common causes C_i of X and Y are such that all directed paths from C_i to X are of the opposite sign as all directed paths from C_i to Y.

It has been shown elsewhere (VanderWeele and Robins, 2006a) that if X and Y are positively monotonically associated then $Cov(E_1, E_2) \geq 0$ and if X and Y are negatively monotonically associated then $Cov(E_1, E_2) \leq 0$. We now develop several results concerning the monotonicity in the conditioning argument of certain conditional expectations.

4. Monotonic Effects and Conditional Expectations

The following lemma can be proved by integration by parts and will be used in the proofs of the subsequent propositions.

Lemma 1. If $h(y, a, r)$ is non-decreasing in y and in a and $S(y|a, r) = pr(Y > y|A = a, R = r)$ is non-decreasing in a for all y then $E[h(Y, A, R)|A = a, R = r]$ is non-decreasing in a.

Proposition 3 immediately follows from Lemma 1.

Proposition 3. Suppose that the $A - Y$ edge, if it exists, is positive. Let X denote some set of non-descendents of Y that includes $\widetilde{pa}_Y$, the parents of Y other than A, then $E[Y|X = x, A = a]$ is non-decreasing in a for all values of x.

Proposition 4 gives the basic result for the monotonicity of conditional expectations. For the conditional expectation of some variable Y to be monotonic in a conditioning argument A, it requires that the conditioning set includes variables that block all backdoor paths A to Y.

Proposition 4. Let X denote some set of non-descendents of A that blocks all backdoor paths from A to Y. Let $R = (R_1, ..., R_m)$ denote an ordered list of some set of nodes on directed paths between A and Y such that for each i the backdoor paths from each element of $R_i, ..., R_m$ to Y are blocked by $R_1, ..., R_{i-1}, A$ and X. If all directed paths between A and Y are positive except possibly through R then $S(y|a, x, r)$ and $E[y|a, x, r]$ are non-decreasing in a.

If $R = \varnothing$ the statement of Proposition 4 is considerably simplified and is stated in the following corollary.

Corollary. Let X denote some set of non-descendents of A that blocks all backdoor paths from A to Y. If all directed paths between A and Y are positive then $S(y|a, x)$ and $E[y|a, x]$ are non-decreasing in a.

Propositions 5-8 relax the condition that the conditioning set includes variables that block all backdoor paths A to Y and impose certain other conditions; the proofs of each of these propositions make use of Proposition 4.

Proposition 5. Suppose that A is not a descendent of Y, that A is binary, and that A and Y are positively monotonically associated then $E[Y|A]$ is non-decreasing in A.

Proposition 6. Suppose that A is not a descendent of Y, that Y is binary, and that A and Y are positively monotonically associated then $E[A|Y]$ is non-decreasing in Y.

Propositions 5 and 6 require that conditioning variable be binary. Counterexamples can be constructed to show that if the conditioning variable is not binary then the conditional expectation may not be non-decreasing in the conditioning argument even if A and Y are positively monotonically associated (see Appendix 3, counterexamples 1 and 2).

Propositions 5 and 6 can be combined to give the following corollary which makes no reference to the ordering of A and Y.

Corollary. Suppose that A is binary and that A and Y are positively monotonically associated then $E[Y|A]$ is non-decreasing in A.

Example 1. Consider the signed directed acyclic graph given in Figure 1.

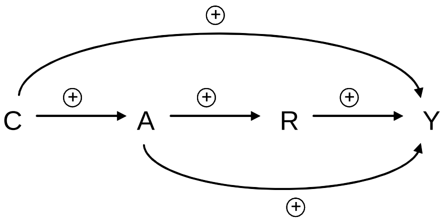

Fig. 1. Example illustrating Propositions 4-6.

By Proposition 4, $E[Y|A=a,C=c,R=r]$ and $E[Y|A=a,C=c]$ are non-decreasing in a since $X=C$ blocks all backdoor paths from A to Y. If A is binary then by Proposition 5, it is also the case that $E[Y|A=a]$ is non-decreasing in a. If Y is binary, then by Proposition 6, $E[A|Y=y]$ is non-decreasing in y.

Propositions 7 and 8 consider the monotonicity of conditional expectations while conditioning on variables other than the variable in which monotonicity holds but not conditioning on variables that are sufficient to block all backdoor paths between A and Y. Propositions 7 and 8 generalize Propositions 5 and 6 respectively.

Proposition 7. Suppose that A is not a descendent of Y and that A is binary. Let Q be some set of ancestors of Y that are not descendents of A and let R be any set of ancestors of Y. If A and Y are positively monotonically associated with the exception that directed paths from A to Y through R need not be of positive sign then $E[Y|A,R,Q]$ is non-decreasing in A.

Proposition 8. Suppose that A is not a descendent of Y and that Y is binary. Let Q be some set of ancestors of Y that are not descendents of A and let R be any set of ancestors of Y. If A and Y are positively monotonically associated with the exception that directed paths from A to Y through R need not be of positive sign then $E[A|Y,R,Q]$ is non-decreasing in Y.

Example 2. Consider the signed directed acyclic graph given in Figure 2.

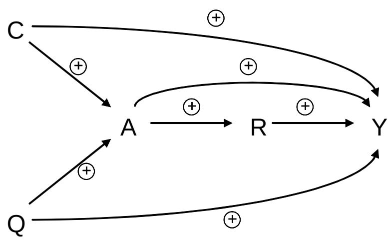

Fig. 2. Example illustrating Propositions 7 and 8.

If A is binary, then by Proposition 7, $E[Y|A=a,C=c,R=r,Q=q]$, $E[Y|A=a,R=r,Q=q]$, $E[Y|A=a,C=c,R=r]$ and $E[Y|A=a,R=r]$ are all non-decreasing in a. If Y is binary then by Proposition 8, $E[A|Y=y,C=c,R=r,Q=q]$, $E[A|Y=y,R=r,Q=q]$, $E[A|Y=y,C=c,R=r]$ and $E[A|Y=y,R=r]$ are all non-decreasing in y.

5. Effect Modification and Monotonic Effects

If conditioning on a particular variable reverses the sign of the effect of another variable on the outcome, then the first variable is said to be a qualitative effect modifier. The following definition gives the condition for qualitative effect modification more formally.

Definition 6. A variable Q is said to be an effect modifier for the causal risk difference of A on Y if Q is not a descendent of A and if there exist two levels of A, a_0 and a_1 say, such that $E[Y_{A=a_1}|Q=q]-E[Y_{A=a_0}|Q=q]$ is not constant in q. Furthermore Q is said to be a qualitative effect modifier if there exist two levels of A, a_0 and a_1, and two levels of Q, q_0 and q_1, such that $sign(E[Y_{A=a_1}|Q=q_1]-E[Y_{A=a_0}|Q=q_1])\neq sign(E[Y_{A=a_1}|Q=q_0]-E[Y_{A=a_0}|Q=q_0])$.

Monotonic effects and weak monotonic effects are closely related to the concept of qualitative effect modification. Essentially, the presence of a monotonic effect precludes the possibility of qualitative effect modification. This is stated precisely in Theorems 1 and 2.

THEOREM 1. Suppose that some parent A_1 of Y is such that the $A_1 - Y$ edge is of positive sign then there can be no other parent, A_2, of Y which is a qualitative effect modifier for causal effect of A_1 on Y, either unconditionally or within some stratum $C = c$ of the parents of Y other than A_1 and A_2.

A similar result clearly holds if the $A_1 - Y$ edge is of negative sign. We give the contrapositive of Theorem 1 as a corollary.

COROLLARY. Suppose that some parent of Y, A_2, is a qualitative effect modifier for causal risk difference of another parent of Y, A_1, either unconditionally or within some stratum $C = c$ of the parents of Y other than A_1 and A_2 then A_1 can have neither a weak positive monotonic effect nor a weak negative monotonic effect on Y.

If there are intermediate variables between A and Y then Theorem 1 can be generalized to give Theorem 2.

THEOREM 2. Suppose that all directed paths from A to Y are of positive sign (or are all of negative sign) then there exists no qualitative effect modifier Q on the directed acyclic graph for the causal effect of A on Y.

EXAMPLE 3. Consider the signed directed acyclic graph given in Figure 3 in which the $A - Y$ edge is of positive sign.

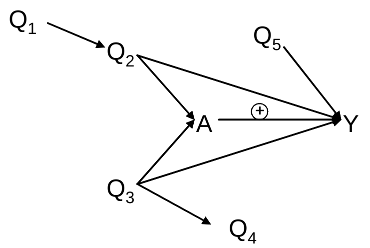

Fig. 3. Example illustrating the use of Theorems 1 and 2.

It can be shown that any of Q_1, Q_2, Q_3, Q_4 or Q_5 can serve as effect modifiers for the causal effect of A on Y (VanderWeele and Robins, 2006d). However, by Theorem 1 or 2, since A has a (weak) monotonic effect on Y, none of Q_1, Q_2, Q_3, Q_4 or Q_5 can serve as *qualitative* effect modifiers for the causal effect of A on Y. Conversely, if it is found that one of Q_1, Q_2, Q_3, Q_4 or Q_5 *is* a qualitative effect modifier for the causal effect of A on Y then the $A - Y$ edge cannot be of positive (or negative) sign.

Appendix 1. Monotonic Effects and Causal Inference.

CAUSAL INFERENCE RESULT 1. If A is an ancestor of Y and the sign of every directed path between A and Y is positive then $E[Y_{A=a}]$ is non-decreasing in a.

CAUSAL INFERENCE RESULT 2. Suppose A is binary variable and an attempt is made to estimate the causal effect of A on Y controlling for a number of variables X which do not block all backdoor paths between A and Y but which do not open any backdoor paths between A and Y which were previously blocked. The true causal effect on Y of intervening to set $A = a$ is given by $E[Y_{A=a}] = \sum_z E[Y|A = a, Z = z]P(Z = z)$ where Z is any set of non-descendents of A blocks all back-door paths from A to Y. The estimate of the causal effect on Y of intervening to set $A = a$ controlling on for X is given by $\sum_x E[Y|A = a, X = x]P(X = x)$. If all unblocked backdoor paths between A and Y given X are of positive sign then $\sum_x E[Y|A = 1, X = x]P(X = x) \geq E[Y_{A=1}]$ and $\sum_x E[Y|A = 0, X = x]P(X = x) \leq E[Y_{A=0}]$. If all unblocked backdoor paths between A and Y are of negative sign then $\sum_x E[Y|A = 1, X = x]P(X = x) \leq E[Y_{A=1}]$ and $\sum_x E[Y|A = 0, X = x]P(X = x) \geq E([Y_{A=0}]$.

Appendix 2. Proofs.

PROOF OF PROPOSITION 1
By the definition of a non-parametric structural equation, $Y_{a,\widetilde{pa}_Y}(\omega) = f(\widetilde{pa}_Y, a, \epsilon_Y(\omega))$ and from this the result follows.

PROOF OF PROPOSITION 2
Since A has a positive monotonic effect on Y, for any $a_1 \geq a_0$ we have that $S(y|a_1, \widetilde{pa}_Y) = P(Y > y|a_1, \widetilde{pa}_Y) = P\{f(\widetilde{pa}_Y, a_1, \epsilon_Y) > y\} \geq P\{f(\widetilde{pa}_Y, a_0, \epsilon_Y) > y\} = P(Y > y|a_0, \widetilde{pa}_Y) = S(y|a_1, \widetilde{pa}_Y)$.

PROOF OF LEMMA 1
For $a \geq a'$ we have $E[h(Y,A,R)|A = a, R = r] - E[h(Y,A,R)|A = a', R = r] = \int_{-\infty}^{\infty} h(y,a,r)dF(y|a,r) - \int_{-\infty}^{\infty} h(y,a',r)dF(y|a',r) = \int_{-\infty}^{\infty} h(y,a,r)d\{F(y|a,r) - F(y|a',r)\} + \int_{-\infty}^{\infty}\{h(y,a,r) - h(y,a',r)\}dF(y|a',r) = [h(y,a,r)\{F(y|a,r) - F(y|a',r)\}]_{-\infty}^{\infty} - \int_{-\infty}^{\infty}\{F(y|a) - F(y|a')\}dh(y;a,r) + \int_{-\infty}^{\infty}\{h(y,a,r) - h(y,a',r)\}dF(y|a',r) = \int_{-\infty}^{\infty}\{S(y|a,r) - S(y|a',r)\}dh(y;a,r) + \int_{-\infty}^{\infty}\{h(y,a,r) - h(y,a',r)\}dF(y|a',r)$. This final expression is non-negative since the integrands of both integrals are non-negative for $a \geq a'$.

PROOF OF PROPOSITION 3
We have that $E[Y|X = x, A = a] = E[Y|\widetilde{pa}_Y, A = a]$ and since A has a (weak) positive monotonic effect on Y, we have that $S(y|a, \widetilde{pa}_Y)$ is non-decreasing in a and it follows from Lemma 1 that $E[Y|X = x, A = a] = E[Y|\widetilde{pa}_Y, A = a]$ is non-decreasing in a.

PROOF OF PROPOSITION 4
Let C denote the set of common causes of A and Y. Let Q denote the set of nodes that are ancestors of A or of Y but are not descendents of A and not common causes of A and Y. Note that if for each i the backdoor paths from each element of $R_i, ..., R_m$ to Y are blocked by $R_1, ..., R_{i-1}, A$ and X they will also be blocked by $R_1, ..., R_{i-1}, A, C$ and Q since any path from R_i to X must pass through some member of $\{A, C, Q\}$. By Theorem 1, considering only those directed paths not blocked by R, we may replace certain variables by their negations so that all edges on all directed paths between A and Y not blocked by R have positive sign. Let $V_1 = A$ and $V_n = Y$ and let $V_1, ..., V_n$ be an ordered list of all the nodes on directed paths between A and Y such that at least one of the directed paths from each node to Y is not blocked by R. Let $\overline{V}_k = \{V_2, ..., V_k\}$. Let Q^k, C^k and R^k be the members of Q, C and R respectively that are ancestors of $\overline{V}_k$. We will show that

$$\begin{aligned} S(v_k|a, v_2, ..., v_{k-1}, c, q, r) &= S(v_k|a, v_2, ..., v_{k-1}, c, q, r^k) \\ &= S(v_k|a, v_2, ..., v_{k-1}, c^k, q^k, r^k) = S(v_k|pa_{v_k}). \end{aligned}$$

All frontdoor paths from R_m to V_k will be blocked given $A, V_2, ..., V_{k-1}, C, Q, R_1, ..., R_{m-1}$ by a descendent R_m which serves as a collider. All backdoor paths from R_m to V_k with an edge going into V_k will be blocked given $A, V_2, ..., V_{k-1}, C, Q, R_1, ..., R_{m-1}$ by pa_{V_k}. All backdoor paths from R_m to V_k with an edge going out from V_k will be blocked given $A, C, Q, R_1, ..., R_{m-1}$by hypothesis for otherwise there would be an a backdoor path from R_m through V_k to Y not blocked by $A, C, Q, R_1, ..., R_{m-1}$. But all backdoor paths from R_m to V_k with an edge going out from V_k which are blocked by $A, C, Q, R_1, ..., R_{m-1}$ will also be blocked by $A, V_2, ..., V_{k-1}, C, Q, R_1, ..., R_{m-1}$ since such a path concluding with an edge going out from V_k which is blocked by $A, C, Q, R_1, ..., R_{m-1}$ but not blocked by $A, V_2, ..., V_{k-1}, C, Q, R_1, ..., R_{m-1}$ would require that one of $V_2, ..., V_{k-1}$ be a collider on the path but then the path would in fact be blocked by the parents of that collider since all the parents of $V_2, ..., V_{k-1}$ are in the set $A, V_2, ..., V_{k-1}, C, Q, R_1, ..., R_{m-1}$. We have thus shown that V_k and R_m are d-separated given $A, V_2, ..., V_{k-1}, C, Q, R_1, ..., R_{m-1}$ and so

$$S(v_k|a, v_2, ..., v_{k-1}, c, q, r) = S(v_k|a, v_2, ..., v_{k-1}, c, q, r_1, ..., r_{m-1}).$$

Similarly, V_k and R_{m-1} are d-separated given $A, V_2, ..., V_{k-1}, C, Q, R_1, ..., R_{m-2}$ and so

$$S(v_k|a, v_2, ..., v_{k-1}, c, q, r_1, ..., r_{m-1}) = S(v_k|a, v_2, ..., v_{k-1}, c, q, r_1, ..., r_{m-2}).$$

We may carry this argument forward to get

$$S(v_k|a, v_2, ..., v_{k-1}, c, q, r) = S(v_k|a, v_2, ..., v_{k-1}, c, q, r^k).$$

All backdoor paths from V_k to $Q\backslash Q^k \bigcup C\backslash C^k$ will be blocked given $A, V_2, ..., V_{k-1}, C^k, Q^k, R^k$ by pa_{v_k}. Since V_k is not a descendent of $Q\backslash Q^k \bigcup C\backslash C^k$ all frontdoor paths from V_k to $Q\backslash Q^k \bigcup C\backslash C^k$ will involve at least one collider which is a descendent of V_k. This collider is not in the conditioning set $A, V_2, ..., V_{k-1}, C^k, Q^k, R^k$ since this entire set precedes V_k and so the collider will block the frontdoor path from V_k to $Q\backslash Q^k \bigcup C\backslash C^k$. Thus V_k and $Q\backslash Q^k \bigcup C\backslash C^k$ are d-separated given $A, V_2, ..., V_{k-1}, C^k, Q^k, R^k$ and so

$$S(v_k|a, v_2, ..., v_{k-1}, c, q, r^k) = S(v_k|a, v_2, ..., v_{k-1}, c^k, q^k, r^k).$$

Furthermore, $A, V_2, ..., V_{k-1}, C^k, Q^k, R^k$ all precede V_k and include all of the parents of V_k and so

$$S(v_k|a, v_2, ..., v_{k-1}, c^k, q^k, r^k) = S(v_k|pa_{v_k}).$$

We have thus shown as desired that

$$\begin{aligned} S(v_k|a, v_2, ..., v_{k-1}, c, q, r) &= S(v_k|a, v_2, ..., v_{k-1}, c, q, r^k) \\ &= S(v_k|a, v_2, ..., v_{k-1}, c^k, q^k, r^k) = S(v_k|pa_{v_k}). \end{aligned}$$

We can express $E\{1(V_n > v)|A, C, Q, R\}$ as

$$E[E[...E[E[1(V_n > v)|A, \overline{V}_{n-1}, C, Q, R]|A, \overline{V}_{n-2}, C, Q, R]|...|A, V_2, C, Q, R]|A, C, Q, R].$$

Now conditional on $A, \overline{V}_{n-1}\backslash V_i, C, Q, R$ we have that

$$E[1(V_n > v)|A, \overline{V}_{n-1}, C, Q, R]$$

is non-decreasing in v_i for $i = 1, ..., n-1$ since V_i has either a weak positive monotonic effect or no effect on V_n. So conditional on $A, \overline{V}_{n-1}\backslash\{V_i, V_{n-1}\}, C, Q, R$ we have that

$$E[1(V_n > v)|A, \overline{V}_{n-1}, C, Q, R]$$

is a non-decreasing function of v_i and v_{n-1} and furthermore,$S(v_{n-1}|a, v_2, ..., v_{n-2}, c, q, r) = S(v_{n-1}|pa_{v_{n-1}})$ is a non-decreasing in v_i for all $a, v_2, ..., v_{i-1}, v_{i+1}, ..., v_{n-2}, c, q, r$ since V_i has either a weak positive monotonic effect or no effect on V_{n-1}. Thus by Lemma 3 we have that conditional on $A, \overline{V}_{n-2}\backslash V_i, C, Q, R,$

$$E[E[1(V_n > v)|A, \overline{V}_{n-1}, C, Q, R]|A, \overline{V}_{n-2}, C, Q, R]$$

is non-decreasing in v_i for $i = 1, ..., n-2$. Carrying the argument forward, conditional on $A, C, Q, R,$ we will have that

$$E[...E[E[1(V_n > v)|A, \overline{V}_{n-1}, C, Q, R]|A, \overline{V}_{n-2}, C, Q, R]|...|A, V_2, C, Q, R]$$

is a non-decreasing function of v_2 and $v_1 = a$ and since A has either a weak positive monotonic effect or no effect on V_2, $S(v_2|a, c, q, r) = S(v_2|pa_{v_2})$ will be non-decreasing in a and thus by Lemma 3,

$$\begin{aligned} &S(y|a, c, q, r) = E[1(V_n > v)|A, C, Q, R] \\ &\quad = E[E[...E[E[1(V_n > v)|A, \overline{V}_{n-1}, C, Q, R]|A, \overline{V}_{n-2}, C, Q, R]|...|A, V_2, C, Q, R]|A, C, Q, R] \end{aligned}$$

will be non-decreasing in a. Now suppose that some set X blocks all backdoor paths from A to Y and that no component of X is a descendent of A then

$$\begin{aligned} S(y|a, x, r) &= E[E[1(Y > y)|a, C, Q, x, r]|a, x, r] \\ &= E[E[1(Y > y)|a, C, Q, r]|a, x, r] = E[E\{1(Y > y)|a, W, r\}|a, x, r] \end{aligned}$$

where W is the subset of C and Q which are either parents of Y or parents of a node on a directed path between A and Y. There can be no unblocked frontdoor paths from A to W given R and X since the nodes in W are not descendents of A and thus any frontdoor path from A to W will be blocked given

R and X either by a collider or by a node in R. All backdoor paths from A to W are blocked given R and X by X since X blocks all backdoor paths from A to Y. From this it follows that all paths from A to W are blocked given R and X and so W is conditionally independent of A given R and X and so we have

$$\begin{aligned} E[E[1(Y>y)|a,W,r]|a,x,r] &= E[E[1(Y>y)|a,W,r]|x,r] \\ &= E[E[1(Y>y)|a,C,Q,r]|x,r]. \end{aligned}$$

Since $E\{1(Y>y)|a,C,Q,r\}$ is non-decreasing in a for all q and c we also have that

$$S(y|a,x,r) = E[1(Y>y)|a,x,r] = E[E[1(Y>y)|a,C,Q,r]|x,r]$$

is non-decreasing in a. Finally, since $S(y|a,x,r)$ is non-decreasing in a, it follows from Lemma 3 that $E[y|a,x,r]$ is also non-decreasing in a.

Proof of Proposition 5
Proposition 5 is in fact a special case of Proposition 7 with $R=\varnothing$ and $Q=\varnothing$. The proof of Proposition 7 is given below.

Proof of Proposition 6
Proposition 6 is in fact a special case of Proposition 8 with $R=\varnothing$ and $Q=\varnothing$. The proof of Proposition 8 is given below.

Proof of Proposition 7
Let C be the common causes of A and Y not in Q. By the law of iterated expectations,

$$\begin{aligned} &E[Y|A=a,R=r,Q=q] \\ &\quad = \sum_c E[Y|A=a,C=c,R=r,Q=q]P(C=c|A=a,R=r,Q=q) \end{aligned}$$

We have by Proposition 4 that $E[Y|A,R,Q,C]$ is non-decreasing in A and in each dimension of C. Also,

$$P(C=c|A=a,R=r,Q=q) = \frac{P(A=a|C=c,R=r,Q=q)P(C=c|R=r,Q=q)}{P(A=a|R=r,Q=q)}$$

and so

$$P(C=c|A=1,R=r,Q=q) = \nu_{r,q}(c)P(C=c|A=0,R=r,Q=q)$$

where

$$\nu_{r,q}(c) = \frac{P(A=0|R=r,Q=q)P(A=1|C=c,R=r,Q=q)}{P(A=1|R=r,Q=q)P(A=0|C=c,R=r,Q=q)}$$

which is non-decreasing in each dimension of c since the numerator is non-decreasing in each dimension of c and the denominator is non-increasing in each dimension of c by Proposition 4. Thus

$$\begin{aligned} &E[Y|A=1,R=r,Q=q] \\ &= \sum_c E[Y|A=1,C=c,R=r,Q=q]P(C=c|A=1,R=r,Q=q) \\ &\geq \sum_c E[Y|A=0,C=c,R=r,Q=q]P(C=c|A=1,R=r,Q=q) \\ &= \sum_c E[Y|A=0,C=c,R=r,Q=q]\nu_{r,q}(c)P(C=c|A=0,R=r,Q=q) \\ &\geq \sum_c E[Y|A=0,C=c,R=r,Q=q]P(C=c|A=0,R=r,Q=q) \\ &= E[Y|A=0,R=r,Q=q]. \end{aligned}$$

The second inequality holds because $E[Y|A=0,R=r,Q=q,C=c]$ is non-decreasing in each dimension of c and $P(C=c|A=1,Q=q,R=r) = \nu_{r,q}(c)P(C=c|A=0,Q=q,R=r)$ weights more heavily higher values of each dimension of c than does $P(C = c|A = 0, Q = q, R = r)$ since $\nu_{r,q}(c)$ is non-decreasing in each dimension of c. Thus $E[Y|A = a, R = r, Q = q]$ is non-decreasing in a.

Proof of Proposition 8

Let C denote the set of common causes of Y and A. The causal directed acyclic graph can be marginalized to one which includes only Y, A and their common causes. Let C denote the set of common causes of Y and A on this causal directed acyclic subgraph then C will include all parents of A on the subgraph. We then have that

$$E[A|Y = y, R = r, Q = q] = \sum_c E[A|Y = y, C = c, R = r, Q = q]P(C = c|Y = y, R = r, Q = q)$$
$$= \sum_{c,a} aP(A = a|Y = y, C = c, R = r, Q = q)P(C = c|Y = y, R = r, Q = q)$$
$$= \sum_{c,a} a\frac{P(Y = y, A = a, C = c|R = r, Q = q)}{P(Y = y, C = c|R = r, Q = q)}P(C = c|Y = y, R = r, Q = q)$$
$$= \sum_{c,a} a\frac{P(Y = y|A = a, C = c, R = r, Q = q)}{P(Y = y|R = r, Q = q)}P(A = a, C = c|R = r, Q = q)$$
$$= E_{C,A}[A\frac{P(Y = y|A, C, R = r, Q = q)}{P(Y = y|R = r, Q = q)}|R = r, Q = q].$$

By Proposition 4 with $X = \{C, Q\}$ we have that conditional on $R = r$ and $Q = q$, $\frac{P(Y=1|A,C,R=r,Q=q)}{P(Y=1|R=r,Q=q)}$ is a non-decreasing function of A and of each dimension of C. Similarly, $\frac{P(Y=0|A,C,R=r,Q=q)}{P(Y=0|R=r,Q=q)}$ is a non-increasing function of A and each dimension of C. Since over c and a, conditional on $R = r$ and $Q = q$, $\frac{P(Y=y|A=a,C=c,R=r,Q=q)}{P(Y=y|R=r,Q=q)}$ is a weight function that sums to 1, we have that

$$\begin{aligned} E[A|Y = 1, R = r, Q = q] &= E_{C,A}[A\frac{P(Y = 1|A, C, R = r, Q = q)}{P(Y = 1|R = r, Q = q)}|R = r, Q = q] \\ &\geq E_{C,A}[A\frac{P(Y = 0|A, C, R = r, Q = q)}{P(Y = 0|R = r, Q = q)}|R = r, Q = q] \\ &= E[A|Y = 0, R = r, Q = q] \end{aligned}$$

and so $E[A|Y, R, Q]$ is non-decreasing in Y.

Proof of Theorem 1

Note that by Proposition 3 above if A_1 has a weak positive monotonic effect on Y then $E[Y|A_1 = a_1, A_2 = a_2, C = c]$ must be non-decreasing in a_1 and if A_1 has a weak negative monotonic effect on Y then $E[Y|A_1 = a_1, A_2 = a_2, C = c]$ must be non-increasing in a_1. Since $(Y \coprod A_1|\{A_2, C\})_{G_{\underline{E_1}}}$ where $G_{\underline{E_1}}$ is the original directed acyclic graph G with all edges emanating from A_1 removed, we have $Y_{A_1=a} \coprod A_1|C$ (Pearl, 1995). Thus $E[Y_{A_1=a_1}|A_2 = a_2, C = c] = E[Y|A_1 = a_1, A_2 = a_2, C = c]$ and so if A_2 is a qualitative effect modifier for the causal effect of A_1 on Y for stratum $C = c$ then we must two values of A_1, a_1^* and a_1^{**}, and two levels of A_2, a_2' and a_2'', such that $E[Y|A_1 = a_1^{**}, A_2 = a_2'', C = c] - E[Y|A_1 = a_1^*, A_2 = a_2'', C = c] < 0$ and $E[Y|A_1 = a_1^{**}, A_2 = a_2', C = c] - E[Y|A_1 = a_1^*, A_2 = a_2', C = c] > 0$. Either $a_1^{**} > a_1^*$ or $a_1^{**} < a_1^*$. Consider the first case (the second is analogous) then since $E[Y|A_1 = a_1^{**}, A_2 = a_2'', C = c] - E[Y|A_1 = a_1^*, A_2 = a_2'', C = c] < 0$, A_1 does not have a weak positive monotonic effect on Y and since $E[Y|A_1 = a_1^{**}, A_2 = a_2', C = c] - E[Y|A_1 = a_1^*, A_2 = a_2', C = c] > 0$, A_1 does not have a weak negative monotonic effect on Y. Now if A_2 is a qualitative effect modifier for the causal effect of A_1 unconditionally then we must have two values of A_1, a_1^* and a_1^{**}, and two levels of A_2, a_2' and a_2'', such that $E[Y_{A_1=a_1^{**}}|A_2 = a_2''] - E[Y_{A_1=a_1^*}|A_2 = a_2''] < 0$ and $E[Y_{A_1=a_1^{**}}|A_2 = a_2'] - E[Y_{A_1=a_1^*}|A_2 = a_2'] > 0$. Once again either $a_1^{**} > a_1^*$ or $a_1^{**} < a_1^*$. We will consider the first case (the second is analogous). We thus have that $\sum_c E[Y|A_1 = a_1^{**}, A_2 = a_2'', C = c]P(C = c|A_2 = a_2'') = \sum_c E[Y_{A_1=a_1^{**}}|A_2 = a_2'', C = c]P(C = c|A_2 = a_2'') = E[Y_{A_1=a_1^{**}}|A_2 = a_2''] < E[Y_{A_1=a_1^*}|A_2 = a_2''] = \sum_c E[Y_{A_1=a_1^*}|A_2 = a_2'', C = c]P(C = c|A_2 = a_2'') = \sum_c E[Y|A_1 = a_1^*, A_2 = a_2'', C = c]P(C = c|A_2 = a_2'')$ and so A_1 cannot have a weak positive monotonic effect on Y and similarly, $\sum_c E[Y|A_1 = a_1^{**}, A_2 = a_2', C = c]P(C = c|A_2 = a_2') = \sum_c E[Y_{A_1=a_1^{**}}|A_2 =$

$a_2', C = c]P(C = c|A_2 = a_2') = E[Y_{A_1=a_1^{**}}|A_2 = a_2'] > E[Y_{A_1=a_1^*}|A_2 = a_2'] = \sum_c E[Y_{A_1=a_1^*}|A_2 = a_2', C = c]P(C = c|A_2 = a_2') = \sum_c E[Y|A_1 = a_1^*, A_2 = a_2', C = c]P(C = c|A_2 = a_2')$ and so A_1 cannot have a weak negative monotonic effect on Y.

Proof of Theorem 2

We prove the Theorem for weak positive monotonic effects. The proof for weak negative monotonic effects is similar. Let C denote all non-descendents of A which are either parents of Y or parents of a node on a directed path between A and Y. By the law of iterated expectations we have $E[Y_{A=a_1}|Q = q] - E[Y_{A=a_0}|Q = q] = \sum_c E[Y_{A=a_1}|C = c, Q = q]P(C = c|Q = q) - \sum_c E[Y_{A=a_0}|C = c, Q = q]P(C = c|Q = q)$. We will show that this latter expression is equal to $\sum_c E[Y_{A=a_1}|C = c]P(C = c|Q = q) - \sum_c E[Y_{A=a_0}|C = c]P(C = c|Q = q)$. By Theorem 3 of Pearl (1995) it suffices to show that $(Y \coprod Q|C, A)_{G_{\overline{A}}}$ where $G_{\overline{A}}$ denotes the graph obtained by deleting from the original directed acyclic graph all arrows pointing into A. Any front door path from Y to Q in $G_{\overline{A}}$ will be blocked by a collider. Any backdoor path from Y to Q in $G_{\overline{A}}$ will be blocked by C. We thus have that $E[Y_{A=a_1}|Q = q] - E[Y_{A=a_0}|Q = q] = \sum_c E[Y_{A=a_1}|C = c]P(C = c|Q = q) - \sum_c E[Y_{A=a_0}|C = c]P(C = c|Q = q)$. Since C will block all backdoor paths from A to Y we have by the backdoor path adjustment theorem $\sum_c E[Y|C = c, A = a_1]P(C = c|Q = q) - \sum_c E[Y|C = c, A = a_0]P(C = c|Q = q) = \sum_c\{E[Y|C = c, A = a_1] - E[Y|C = c, A = a_0]\}P(C = c|Q = q)$. If there were a qualitative effect modifier Q for the causal effect of A on Y then there would exist a value q_0 such that $E[Y_{A=a_1}|Q = q_0] - E[Y_{A=a_0}|Q = q_0] < 0$. But since all paths between A and Y are of positive sign and since C blocks all backdoor paths from A to Y we have by Proposition 4 that $E[Y|C = c, A = a]$ is non-decreasing in a and so $E[Y_{A=a_1}|Q = q_0] - E[Y_{A=a_0}|Q = q_0] = \sum_c\{E[Y|C = c, A = a_1] - E[Y|C = c, A = a_0]\}P(C = c|Q = q_0) \geq 0$.

Appendix 3. Counterexamples.

Counterexample 1

Consider the directed acyclic graph given in Figure 4.

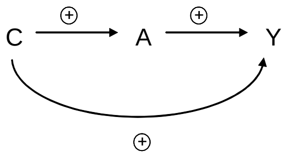

Fig. 4. Directed acyclic graph illustrating counterexamples to Propositions 5 and 6 when A is not binary.

In this example C and Y are binary and A is ternary. Suppose that $C \sim Ber(0.5)$, $\epsilon_A \sim Ber(0.5)$ and that if $\epsilon_A = 0$ then $A = 0$ and if $\epsilon_A = 1$ then $A = C + 1$. Suppose also that $A = 2$ then $Y = 1$ and that if $A = 0$ or $A = 1$ then $Y = C$. Clearly then C has a positive monotonic effect on A and on Y and A has a positive monotonic effect on Y and so A and Y are positively monotonically associated. However, we have that $E[Y|A = 1] = E[C|A = 1] = 0 * P(C = 1|A = 1) = 0$ but $E[Y|A = 0] = E[C|A = 0] = 1 * P(C = 1|A = 0) + 0 * P(C = 0|A = 0) = 1/2$.

Counterexample 2

Consider again the directed acyclic graph given in Figure 4. In this example we will assume that C and A are binary and that Y is ternary. Suppose that $C \sim Ber(0.5)$ and that ϵ_A takes on the values 0, 1 and 2, each with probability 1/3. Suppose also that if $\epsilon_A = 0$ then $A = 0$, if $\epsilon_A = 1$ then $A = C$ and if $\epsilon_A = 2$ then $A = 1$. Suppose further that if $C = 0$ then $Y = 0$ and if $C = 1$ then $Y = A + 1$. Clearly then C has a positive monotonic effect on A and on Y and A has a positive monotonic effect on Y and so A and Y are positively monotonically associated. However, we have that $E[A|Y = 1] = 0$ but $E[A|Y = 0] = E[A|C = 0] = 1/3$.

References

Dawid, A.P. (2002). Influence diagrams for causal modelling and inference. *Int. Statist. Rev.* **70**, 161-189.

Geiger, D., Verma, T.S. and Pearl, J. (1990). Identifying independence in Bayesian networks. *Networks* **20**, 507-534.

Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations. *Econometrica* **11**, 1-12.

Hardle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (1998). *Wavelets, Approximation, and Statistical Applications.* New York: Springer-Verlag.

Lauritzen, S.L., Dawid, A.P., Larsen, B.N. and Leimer, H.G. (1990). Independence properties of directed Markov fields. *Networks* **20**, 491-505.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika* **82**, 669-688.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference.* Cambridge: Cambridge University Press.

Robins, J. M. (2003). Semantics of causal DAG models and the identification of direct and indirect effects. In *Highly Structured Stochastic Systems* (eds. P. Green, N.L. Hjort, and S. Richardson), 70-81. New York: Oxford University Press.

Spirtes, P., Glymour, C. and Scheines, R. (1993). *Causation, Prediction and Search.* New York: Springer-Verlag.

VanderWeele, T.J. and Robins, J.M. (2006a). Signed directed acyclic graphs for causal inference. *Submitted.*

VanderWeele, T.J. and Robins, J.M. (2006b). Minimal sufficient causation and directed acyclic graphs. *Submitted.*

VanderWeele, T.J. and Robins, J.M. (2006c). Biologic interactions and their identification. *Submitted.*

VanderWeele, T.J. and Robins, J.M. (2006d). A structural classification of effect modification. *Submitted.*

Verma, T. and Pearl, J. (1988). Causal networks: Semantics and expressiveness. In *Proceedings of the 4th Workshop on Uncertainty in Artificial Intelligence*, 352-359. Reprinted in *Uncertainty in Artificial Intelligence* (ed. R. Shachter, T.S. Levitt, and L.N. Kanal), **4**, 69-76. Amsterdam: Elesevier.

Wellman, M.P. (1990). Fundamental concepts of qualitative probabilistic networks. *Artificial Intelligence* **44**, 257-303.

Wright, S. (1921). Correlation and causation. *J. Agric. Res.* **20**, 557-585.