Two central assumptions concerning causal inference in the potential outcomes framework for estimating neighborhood effects are examined. The stable unit treatment value assumption in the context of neighborhood effects requires that an individual’s outcome does not depend on the treatment assigned to neighborhoods other than the individual’s own neighborhood. The assumption is important in order make estimation feasible and some discussion is given concerning the contexts in which the neighborhood-level stable unit treatment value assumption is likely to hold. The ignorability assumption allows the researcher to move from conclusions about association to conclusions about causation. In the context of neighborhood-wide interventions, the ignorability assumption for the individual-level potential outcomes framework can be easily adapted for neighborhood effects.

*Keywords*: Causal inference; ignorability; multilevel models; neighborhood effects; potential outcomes.
1. Introduction

Several different types of inquiry arise when considering questions of "neighborhood effects." We might be interested in how health or income outcomes would differ if individual $i$ were moved from neighborhood $j$ to neighborhood $k$. Alternatively we might be interested in how individual or population outcomes would likely change under a particular neighborhood-level intervention. We could consider, for example, the effect of building a new hospital, or of increasing the police force in a particular area or of fixing local roads and sidewalks. In some studies, researchers are able to randomize interventions to neighborhoods (e.g. [1]) or sometimes are even able to randomize individuals to a change in neighborhood (e.g. [2]). Studies of the former type are generally referred to as randomized community trials; such studies are well understood and have been employed with some frequency [3]. With the Moving to Opportunity (MTO) study sponsored by the U.S. Department of Housing and Urban Development, there has been more recent interest in the interpretation of studies which randomize individuals to a change in neighborhood [2, 4, 5]. In many studies, however, in which the aim is to estimate neighborhood effects, observational data is used. This raises questions about the assumptions that are required to move from associational to causal conclusions in such observational settings. The potential outcomes framework has been routinely employed to articulate these assumptions in the context of non-clustered cohorts [6, 7]. However, the manner in which the concepts from the potential outcomes framework extend to the study of neighborhood effects in which individuals are clustered within neighborhoods is not well documented or understood. In this paper our focus will be the effects of a neighborhood-level intervention rather than the effects of moving an individual from one neighborhood to another. We will consider both randomized community trials and observational data settings. We will examine the technical assumptions need to draw causal conclusions about neighborhood effects from such data.

2. Multilevel Models and Neighborhood Effects

Multilevel modeling (also called hierarchical modeling) is often employed in the estimation of the causal effects of neighborhood-level interventions or characteristics. Let $Y_{ik}$ denote the outcome for individual $i$ in neighborhood $k$. Let $X_{ik}$ denote individual-level characteristics for for individual $i$ in neighborhood $k$. Let $Z_k$ denote neighborhood-level characteristics for neighborhood $k$. Let $T_k$ denote the neighborhood wide treatment or intervention under consideration. The typical model employed in neighborhood effects research concerning a neighborhood-level
intervention will have the following form.

\[ Y_{ik} = \alpha + \beta_1 X_{ik} + \beta_2 Z_k + \gamma T_k + u_k + \epsilon_{ik} \]  

(1)

where \( \alpha \) denotes an intercept, \( u_k \) denotes a neighborhood-level random error and \( \epsilon_{ik} \) denotes an individual-level random error. The \( u_k \) and \( \epsilon_{ik} \) are usually assumed to be independent and identically distributed with mean zero. That the \( \epsilon_{ik} \) 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 (1) is also often written in hierarchical form with level 1 given by the equation:

\[ Y_{ik} = \mu_k + \beta_1 X_{ik} + \epsilon_{ik} \]  

(2)

and level 2 given by the equation:

\[ \mu_k = \alpha + \beta_2 Z_k + \beta_3 T_k + u_k. \]  

(3)

Substituting equation (3) into equation (2) gives equation (1). Equation (1) could of course also be generalized to include interaction terms. Estimation of the coefficients of (1) or equivalently of (2) and (3) is straightforward using standard statistical software. Several text books cover such estimation procedures; one such textbook [8] provides some discussion of causal inference with the multilevel model.

For purposes of illustration 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 age, sex, race, socioeconomic status, and prior health status, all measured at the beginning of year \( r \). Let \( Z_k \) denote neighborhood-level characteristics for neighborhood \( k \) including number of existing hospitals, neighborhood safety and neighborhood mean income, also measured at the beginning of year \( r \).

Before proceeding we note that not all neighborhood-level interventions or randomized community trials can be classified as "neighborhood effects research." Consider, for example, a randomized community trial in which all households in the neighborhoods selected for "treatment" are mailed a nine month supply of vitamin supplements. Such a study would probably not be considered "neighborhood research" because the intervention is such that the neighborhood itself is not in any
way integrally altered. The line between what types of interventions do and do not constitute "neighborhood research" may not always be entirely clear - one might imagine a scenario in which the vitamin supplement mailings lead to greater community discussion of nutrition and consequently a more "health-conscious community." Nevertheless there is a considerable difference between vitamin supplement mailings and an intervention that intentionally tries to create a more health-conscious community by mailing health information, promoting community discussion, making use of billboards and local newspapers, etc.

**Ignorability and Stability Assumptions**

In this section we consider the assumptions necessary to provide a causal interpretation of the coefficient $\gamma$. We first consider the neighborhood-level stable unit treatment value assumption (NL-SUTVA). The neighborhood-level stable unit treatment value assumption requires that an individual’s outcome does not depend on the treatment assigned to neighborhoods other than the individual’s own neighborhood. Let $Y_{ik}(T_k = t)$ denote the outcome individual $i$ in neighborhood $k$ would have obtained if the neighborhood-level treatment $T_k$ in neighborhood $k$ were set to $t$. The potential outcome $Y_{ik}(T_k = t)$ is only well defined under NL-SUTVA. If NL-SUTVA does not hold then the individual’s outcome may depend on the treatments assigned to neighborhoods other than the individual’s own neighborhood and so the quantity we wish to define by $Y_{ik}(T_k = t)$ may take on many different values depending upon the treatment status of neighborhoods other than neighborhood $k$. Under NL-SUTVA, $Y_{ik}(T_k = t)$ is uniquely defined. Note that NL-SUTVA does not require that there be no treatment interaction between two individuals in the same neighborhood; rather NL-SUTVA requires that there be no treatment interaction between individuals in different neighborhoods, or once again more generally, that the treatment received by one neighborhood does not affect outcomes for individuals in other neighborhoods. Consider, for example, a neighborhood crime prevention program that affects members of two different gangs in the same neighborhood in such a way that a particular fight between two individuals is prevented only if both gang members are exposed to the crime prevention program; if only one were exposed the fight would occur. In this case, there is interaction or "interference" between the two individuals but no interference from treatments assigned to different neighborhoods. Although the individual-level SUTVA condition is not satisfied, NL-SUTVA is satisfied so long as both gangs operate in the same neighborhood.

Clarifying the stable unit treatment value assumption for the neighborhood effects context may have important implications for study design. Although the researcher
has no control over treatment assignment in observational studies, in both observational studies and randomized trials the researcher will likely have considerable choice in which experimental units to include in the study. As discussed above, NL-SUTVA will be violated if the outcome of individual $i$ in neighborhood $k$ is affected by treatment in neighborhood $j \neq k$. Thus, if an individual’s health outcomes are affected by easier access to health care from the construction of a new hospital in a nearby neighborhood, NL-SUTVA will be violated. A researcher designing a randomized trial might thus choose the experimental units, the neighborhoods, in such a way that there would be considerable distance between the neighborhoods so as to prevent the possibility of the construction of a new hospital in neighborhood $j$ from affecting individuals in any other neighborhood $k \neq j$ i.e. by ensuring that neighborhoods near neighborhood $j$ are not in the study (cf. [9]). If in an observational context NL-SUTVA is violated then it will be necessary to use techniques which relax this assumption and allow for the modeling of spillover effects [10, 11]; see [5] for the relaxation of SUTVA conditions in a setting in which individuals are randomized to a change of neighborhoods. This will of course be a necessity if spillover effects (i.e. between-group interactions) are in fact of primary interest.

We next turn to the second central assumption needed to draw causal conclusions from statistical models like that given in equation (1). The ignorability assumption makes certain conditional independence claims about the counterfactual outcomes and the treatment and thereby allows the researcher to move from conclusions about association to conclusions about causation. Intuitively, the ignorability assumption is often understood as indicating that within strata of certain measured, potentially confounding variables, the treatment is essentially as though it were randomized. More formally, we will say that neighborhood treatment assignment is ignorable given the covariates $X_{ik}$ and $Z_k$ if for $t = 0, 1$

$$Y_{ik}(T_k = t) \prod T_k | X_{ik}, Z_k$$

(4)

where $\prod$ denotes the independence relation. The ignorability assumption states that conditional on the covariates $X_{ik}, Z_k$ the potential outcomes $Y_{ik}(T_k = 0)$ and $Y_{ik}(T_k = 1)$ are independent of treatment $T_k$. This ignorability condition will hold if treatment is randomized within strata of $X_{ik}$ and $Z_k$. Condition (4) is in fact somewhat weaker than treatment randomization within strata of $X_{ik}$ and $Z_k$. Treatment randomization within strata of $X_{ik}$ and $Z_k$ implies that treatment $T_k$ is conditionally independent not only of the potential outcomes $Y_{ik}(T_k = 0)$ and $Y_{ik}(T_k = 1)$ given $X_{ik}$ and $Z_k$ but also of all other variables. Condition (4) requires only that
treatment $T_k$ be conditionally independent of the potential outcomes given $X_{ik}$ and $Z_k$. Intuitively, the ignorability assumption (4) is often interpreted as requiring that for every neighborhood $k$ and for each individual $i$ the covariates $X_{ik}, Z_k$ include all relevant variables that affect both the neighborhood-level treatment $T_k$ and the potential outcomes $Y_{ik}(T_k = 0)$ and $Y_{ik}(T_k = 1)$ i.e. that $X_{ik}$ and $Z_k$ contain all the individual and neighborhood level confounding variables. Thus the covariate vector $Z_k$ would likely need to include the number of previously existing hospitals in the neighborhood, the presence of hospitals in adjacent neighborhoods along with perhaps the mean income for the neighborhood and any other neighborhood-level characteristics that affect both the construction of a new hospital and individual health outcomes. The situation with the individual-level covariates is more subtle. An individual-level covariate needs to 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 $Z_k$. For example, if the income distribution of the neighborhood were used in making decisions about the construction of a new hospital and if an individual’s income affected health outcomes then the covariate vector $X_{ik}$ would have to include the income for individual $i$ in neighborhood $k$ because each individual’s income affects both the individual’s health outcomes and the overall income distribution which affects decisions about hospital construction. However, if only mean income were used in making decisions about the construction of a new hospital, rather than the income distribution, then to ensure ignorability condition (4) it would suffice to include mean income in the covariate vector $Z_k$; no individual-level income would need to be included among the covariates. Nevertheless, as discussed below, the individual-level income data may be included in the statistical model in order to improve efficiency.

In practice, finding covariates $X_{ik}$ and $Z_k$ such that the treatment ignorability condition (4) will hold may be challenging. Many variables that may affect both treatment and the outcome, such as for example political resourcefulness, trust and community solidarity, will often be very difficult to measure. Condition (4) will at best hold only approximately and a central challenge of the study’s design will be collecting sufficiently rich data so that the approximation will be a reasonable one. Of course, as indicated above, in a community randomized trial the ignorability condition holds by design.

We need one final assumption which is usually referred to as "consistency." The consistency assumption states that when treatment $T_k = t$ then the counterfactual outcome intervening to set treatment $T_k$ to $t$ is equal to the observed outcome i.e. $T_k = t \implies Y_{ik}(T_k = t) = Y_{ik}$. The assumption is necessary mathematically but is
intuitively obvious 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. With NL-SUTVA and the consistency assumption, the neighborhood treatment ignorability assumption allows us to derive estimates of causal effect of treatment from observational data as stated in the following proposition.

Proposition. If the neighborhood treatment assignment is ignorable given $X_{ik}$ and $Z_k$ then

\[
E[Y_{ik}(T_k = t)] = \sum_{x,z} E[Y_{ik}|T_k = t, X_{ik} = x, Z_k = z]P(X_{ik} = x, Z_k = z) \tag{5}
\]

where the expectation is taken over all individuals and neighborhoods. Furthermore under model (1) we have that $E[Y_{ik}(T_k = 1)] - E[Y_{ik}(T_k = 0)] = \gamma$.

Proof. By the law of iterated expectations we have that,

\[
E[Y_{ik}(T_k = t)] = \sum_{x,z} E[Y_{ik}(T_k = t)|X_{ik} = x, Z_k = z]P(X_{ik} = x, Z_k = z)
\]

\[
= \sum_{x,z} E[Y_{ik}|T_k = t, X_{ik} = x, Z_k = z]P(X_{ik} = x, Z_k = z) \text{ by ignorability}
\]

\[
= \sum_{x,z} E[Y_{ik}|T_k = t, X_{ik} = x, Z_k = z]P(X_{ik} = x, Z_k = z) \text{ by consistency.}
\]

Furthermore, if model (1) holds then,

\[
E[Y_{ik}(T_k = 1)] - E[Y_{ik}(T_k = 0)]
\]

\[
= \sum_{x,z} \{E[Y_{ik}|T_k = 1, X_{ik} = x, Z_k = z] - E[Y_{ik}|T_k = 0, X_{ik} = x, Z_k = z]\}P(X_{ik} = x, Z_k = z)
\]

\[
= \sum_{x,z} \gamma P(X_{ik} = x, Z_k = z) = \gamma.
\]

Thus, in our example, $\gamma$ represents the average causal effect of building a new hospital on the health of individuals in the neighborhood in which the hospital is built. The left hand side of equation (5) is a causal quantity. The expression $E[Y_{ik}(T_k = t)]$ denotes the expected value of the counterfactual outcome $Y$ for an individual living in a neighborhood in which there is an intervention to set treatment $T_k$ to $t$. The right hand side of equation (5) can be estimated from the data. The expression $E[Y_{ik}|T_k = t, X_{ik} = x, Z_k = z]$ is in fact the conditional expectation that
is estimated in the multilevel model in equation (1). Of course, for the estimates obtained from the multi-level model to be valid, the functional form of the model must be correctly specified. Thus quadratic terms, interaction terms and cross-level interaction terms between covariates in $X_{ik}$ and covariates in $Z_k$ may need to be included in the model. Appropriate multilevel modeling techniques must be used to estimate the variance of the $E[Y_{ik}(T_k = t)|X_{ik} = x, Z_k = z]$ estimates. If the effect of treatment on the study sample is of interest then $P(X_{ik} = x, Z_k = z)$ can be taken to be fixed weights corresponding to the relative frequency of the covariates $X_{ik}$ and $Z_k$ in the study population. If the study sample consists of a simple random sample of subjects from a specific study population then the variances of the estimates of the effect of treatment on the study population can be computed by bootstrapping techniques.

4. Discussion

In this final section we provide some discussion concerning the estimation and causal interpretation of multilevel model coefficients in neighborhood effects research along with some discussion on the limitations of and possible extensions to these methods. As noted above, for the coefficient estimate $\gamma$ to be interpreted causally the covariate vectors $X_{ik}$ and $Z_k$ must be chosen so that the ignorability condition (4) holds i.e. $X_{ik}$ and $Z_k$ must include all variables that affect both the treatment and the outcome. If some variable in $X_{ik}$ or $Z_k$ affects only the outcome and not the treatment then including the variable in model (1) may increase the efficiency of the coefficient estimates. Provided such variables are measured prior to treatment the ignorability condition (4) will in general be unaffected by their inclusion in the covariates vectors $X_{ik}$ and $Z_k$; exceptions can, however, occur [12, 13]. Graphical models can be useful in assessing the conditional independence assumption required by treatment ignorability [12, 14, 15]. Glymour provides some discussion specifically on the use of graphical models to assess common problems in social epidemiology [16] which may be of interest to those doing neighborhood effects research.

Some comments are also in order concerning the modeling assumptions in the estimation of the multilevel model given in equation (1). Equation (1) assumes a linear relationship between the covariates in the model and the outcome. This assumption will no doubt at best be an approximation to the actual functional relationship between the outcome and the covariates. However, in order to assess whether such a linearity assumption is reasonable and in order to genuinely separate the treatment effect from the effect of the covariates, it is important that there is considerable overlap between the distribution of the individual and neighborhood-
level covariate values in the neighborhoods receiving treatment \((T_k = 1)\) and the neighborhoods which are controls \((T_k = 0)\). Otherwise, significant and potentially scientifically meaningful estimates of the treatment effect may in fact be an artifact of the linearity assumption. Propensity score methods can be of considerable use in assessing these issues concerning the comparability of the treatment and control group [17-19]. Oakes [20] has pointed out how these comparability assumptions can be particularly problematic in neighborhood effects research which attempts to estimate the effects of neighborhood-level socioeconomic status on say health status. This is because the effect of neighborhood-level socioeconomic status is confounded by individual-level socioeconomic status and although individual-level income data can be controlled for in a model there will often be insufficient overlap in the distribution of individual-level income between rich and poor neighborhood. That is to say, the wealthy tend to live in wealthy neighborhoods and the poor in poor neighborhoods and so it is difficult to separate the effects of individual socioeconomic status from neighborhood-level socioeconomic status. Such considerations can sometimes be partially remedied by appropriate study design in which neighborhoods of only modest difference in socioeconomic status are compared.

The analyses we have been considering need to be distinguished from studies in which treatment occurs at the individual-level (so that not all individuals in a particular neighborhood receive treatment) but in which only aggregate level data is available. Within the epidemiologic literature these studies are known as "ecological studies." In such cases, neighborhood-level outcomes are modeled as functions of neighborhood-level characteristics. It has long been recognized [21, 22] that associations at the neighborhood or neighborhood-level could differ considerably, or even be in the opposite direction, from the corresponding associations at the individual-level. Greenland and Morgenstern [23] documented how even in the absence of bias within groups due to confounding, selection, misclassification, etc., neighborhood associations may not reflect individual association if either (i) the outcome in the untreated group varies between groups or (ii) the treatment effect is not additive between groups [cf. 24]. Furthermore, variables may be confounders even if they are unassociated with treatment in every group, especially if they are associated with treatment across groups [23, 25]. Moreover, control for such variables may not reduce the bias produced by these confounding variables [23, 26]. Greenland [27] has also noted that if both a neighborhood characteristic and its individual-level equivalent have an effect on the outcomes then even if neither conditions (i) or (ii) above apply, estimates from ecologic studies are not purely neighborhood effects but represent a combination of neighborhood and individual-level effects. Interpretation becomes
even more difficult in the case of non-linear models. Haneuse and Wakefield [28, 29] have done some work on inference in ecological studies in which the ecologic data are supplemented with a sample of individual-level case control data.

The discussion in this paper is easily generalized to non-continuous outcomes through the use of generalized linear models and also to cases with categorical or continuous, rather than binary, treatments. The stable unit treatment value assumption discussed in Section 2 remains the same under these more general settings. With continuous treatment, the ignorability assumption given in (4) must be modified so as to apply not only to $t = 0, 1$ but to any value of $t$ in the support of $T_k$. For generalized linear multilevel models, equation (1) or equivalently equations (2) and (3) must be modified and appropriate multilevel estimation techniques employed but the causal assumptions required for the valid estimation of neighborhood effects remain essentially the same. Multilevel methods thus do indeed hold some promise for neighborhood effects research. The conclusions drawn from observational data are always tentative and some of the assumptions required may be difficult to meet in practice; however, a firm understanding of the assumptions required for a causal interpretation of estimates from statistical models allows the researcher to critically assess the likely validity of study conclusions.

References


