IFAU
Institute for Evaluation of Labour
Market and Education Policy

On statistical methods for labor market evaluation under interference between units

Maria Karlsson
Mathias Lundin

WORKING PAPER 2016:24

The Institute for Evaluation of Labour Market and Education Policy (IFAU) is a research institute under the Swedish Ministry of Employment, situated in Uppsala. IFAU's objective is to promote, support and carry out scientific evaluations. The assignment includes: the effects of labour market and educational policies, studies of the functioning of the labour market and the labour market effects of social insurance policies. IFAU shall also disseminate its results so that they become accessible to different interested parties in Sweden and abroad.

Papers published in the Working Paper Series should, according to the IFAU policy, have been discussed at seminars held at IFAU and at least one other academic forum, and have been read by one external and one internal referee. They need not, however, have undergone the standard scrutiny for publication in a scientific journal. The purpose of the Working Paper Series is to provide a factual basis for public policy and the public policy discussion.

More information about IFAU and the institute's publications can be found on the website www.ifau.se

ISSN 1651-1166

On statistical methods for labor market evaluation under interference between units

An overview with discussion

Maria Karlsson,
Department of Statistics, USBE, Umeå University

and

Mathias Lundin
Department of Statistics, USBE, Umeå University

Abstract

Evaluation studies aim to provide answers to important questions like: How does this program or policy intervention affect the outcome variables of interest? In order to answer such questions, using the traditional statistical evaluation (or causal inference) methods, some conditions must be satisfied. One requirement is that the outcomes of individuals are not affected by the treatment given to other individuals, i.e., that the no-interference assumption is satisfied. This assumption might, in many situations, not be plausible. However, recent progress in the research field has provided us with statistical methods for causal inference even under interference. In this paper, we review some of the most important contributions made. We also discuss how we think these methods can or cannot be used within the field of policy evaluation and if there are some measures to be taken when planning an evaluation study in order to be able to use a particular method. In addition, we give examples on how interference has been dealt with in some evaluation applications including, but not limited to, labor market evaluations, in the recent past.

Keywords: causal effect, causal inference, contagion effect, direct and indirect effects, evaluation studies, neighborhood effect, peer effect, peer influence effect, policy intervention, spillover effect, SUTVA, treatment effect.

1 Introduction

Empirical evaluation studies of labor market programs, e.g., education programs or employment subsidies, aim to estimate the (causal) effect of the program on some outcome variables of interest, e.g., employment status, time to employment or future earnings. The effect is often called a *treatment effect* since the participation in a program is considered as taking/receiving a treatment.

Related to labor market program evaluations are evaluation studies of policy interventions[1] within other areas, e.g., evaluation of the effects of social insurance policy interventions or of education policy interventions. Hence, most of this paper is applicable also to such evaluations studies and some examples from these areas will occur in the paper as well.

Increased availability of administrative data on individual level (i.e., micro data) has been a driving force for development of methods within the area of empirical labor market program evaluation (van der Klaauw, 2014).

Rubin's model for causal inference (Rubin, 1974) is one of the most popular frameworks for program evaluation. An important assumption in Rubin's model is the *no-interference* assumption saying that the outcomes of individuals are not affected by the treatment given to other individuals. However, this assumption may, in many situations in practice, not be a plausible assumption; neither in randomized experiments nor in non-randomized studies. For example, if a labor market program has an effect on one individual receiving treatment, it could also affect other individuals in the same local labor market due to congestion (Heckman et al., 1999; Rubin, 2010; Gautier et al., 2015). Also, if many individuals are treated, i.e., that the number of trained qualified job seekers increases, there may be more vacancies posted by employers (Ferracci et al., 2014; Gautier et al., 2015).

This means that, even if the treatment would be randomized to individuals, it would not be a good idea to estimate an average treatment effect by comparing average outcomes of treated with the average outcomes of untreated individuals, since the outcomes of the untreated are also altered by the treatment. Instead the direct comparison between treated and untreated should be supplemented with measures of *indirect* effects, which is also called *spillover*, *peer*, *contagion*, or *neighborhood* effects, depending on the context.

It is only in the last few years that the literature on causal inference in the presence of interference has begun to grow and is now a rapidly growing

[1] The terms policies and programmes will be used interchangeably in this paper.

area. In this paper, we review some of the most important contributions made regarding statistical methods[2] for causal inference under interference. The aim is to provide an accessible description of these methods targeted at a wider audience, including policy makers and practitioners. Evaluation of program outcomes is essential for successful policy development and not taking interference into account could result in incorrect conclusions.

The paper is organized as follows: in Section 2, the concept of interference is further described while the suggested methods for causal inference under interference are presented in Section 3. In Section 4 empirical examples of recent evaluations studies with interference are given. The paper concludes with a discussion, where we summarize and discuss how we think that the reviewed methods for causal inference under interference can or cannot be used within the field of labor market evaluations and other policy intervention evaluation studies.

2 Interference

In Rubin's model for causal inference (Rubin, 1974) an individual has one potential outcome for each type of treatment available. In a common setting there are two available treatments, $z = 0, 1$, (e.g., a control treatment and an active treatment) and each individual therefore has two potential outcomes, denoted $Y_j(0)$ and $Y_j(1)$ for individual j. The causal effect of the active treatment versus the control treatment is the difference between the two potential outcomes.

The fundamental problem of causal inference is that only one of the potential outcomes is observed for each individual and thus the causal effect cannot be measured at the individual level. On group level the average causal effect can be estimated by comparing outcomes for the two treatment groups. If treatment assignment is randomized, the *average causal effect* can be estimated by the difference between the mean outcomes in the two groups.

In a non-randomized setting there may be systematic differences in potential outcomes of the groups, e.g, those who choose one of the treatments may on average have higher potential outcomes than those who choose the other treatment, which could lead to biased estimates if we naively compare the observed outcomes of the two groups. Therefore, to make the groups comparable, it is necessary to adjust for confounding variables in the esti-

[2]It should be noted that this means that our purpose is not to also review all important method proposals within related research areas such as econometrics, biometrics, and psychometrics.

mation of the causal effect. This can, for example, be done parametrically through regression methods or non-parametrically through matching methods.

A crucial assumption in Rubin's model is SUTVA, the stable unit treatment value assumption (Rubin, 1980), which consists of two parts. First, a treatment is homogeneous in the sense that there are *not different versions* of the treatment. Second, each individual's outcome depends on the treatment he/she receives regardless of which treatment any other individual receives. If the latter is not the case, there is *interference* between individuals. This paper addresses situations where the second part of SUTVA, the no-interference assumption, is not plausible.

If the no-interference assumption is not satisfied, an individual does not have one potential outcome for each type of treatment. Instead an individual has one potential outcome for each possible combination of treatment statuses amongst all n individuals, i.e., 2^n potential outcomes[3], if there are two treatments available. Comparing the treatment groups as described above can then result in misleading conclusions about the effect of a treatment.

According to Ogburn and VanderWeele (2014) interference can be divided into "three distinct causal pathways by which one individual's treatment may affect another's outcome". Under *direct interference* one individual's treatment directly influences another individual's outcome. An example of this form of interference could be in an educational situation where a treated later exposes an untreated to the knowledge acquired during the course. If we compare those who took the course with those who did not, we might estimate a very small effect on knowledge from the treatment because of spillover of knowledge from treated to untreated individuals.

Interference by contagion is an indirect form of interference which goes from a treated individuals's treatment via his/her outcome to another individual's outcome. A vaccinated individual may avoid getting a disease and thereby not passing it on to another individual. If we compare infection rates between vaccinated and unvaccinated individuals, the effect of the vaccination might seem small. However, vaccinated individuals avoid getting the disease and thereby not exposing unvaccinated individuals and

[3]Note that n^2 is the maximum possible number of (unique) potential outcomes under interference in the scenario with two treatments. The nature of interference determines how many (unique) potential outcomes there are, e.g., there are situations where only the number of treated individuals matters (and not which particular individuals that are treated). Then, the number of (unique) potential outcomes is smaller than n^2, since potential outcomes are the same for several combinations o treatment stuauses amongst the individuals.

thus the "true" effect of the treatment could be much larger than what we have estimated.

The third and most complex form is called *Allocational interference* under which treatment allocates individuals into groups where the group composition affects individual outcomes. An example of where this kind of interference might be present is allocation of children to classes in schools or preschools. A child's achievement in school might be affected by the composition of his/her class.

3 Methods for causal inference under interference

In recent years, the literature suggesting methods for causal inference in the presence of interference has begun to grow. Pioneering work include Sobel (2006), Rosenbaum (2007), and Hudgens and Halloran (2008) and earlier, in the context of vaccine studies, Halloran and Struchiner (1991) and Halloran and Struchiner (1995). These, and most of the more recent proposals, are methods for randomized studies but there are also some suggestions for non-randomized studies.

In this section, we present an overview of proposals. The overview is organized based on how the data is structured. The data structures considered in this overview are *clustered data* (Section 3.1) and *network data* (Section 3.2). Moreover, we distinguish between the methods suggested for randomized experiments and those suggested for non-randomized studies, i.e., observational studies.

VanderWeele et al. (2014) also present a review of literature on causal inference under interference, but most of their summary focuses on methods for randomized experiments. As a consequence, our overview will partly overlap with their survey. The most obvious overlap being in Section 3.1.1.

It is also worth noting that there are other proposals, which we do not cover in this paper, e.g., when data are "paired", i.e., clusters of size two. For example, studies where one of two individuals in a household get treated and the other does not. The interested reader is referred to the following papers: Chiba (2012), Halloran (2012), Halloran and Hudgens (2012), and VanderWeele et al. (2012).

3.1 Clustered data

Many of the proposals on causal inference in the presence of interference are for situations where the individuals under study are clustered and where one allows interference between individuals within a cluster (or group) but

not between clusters. The idea is that the clusters should be separated in some way so that interference between individuals in different clusters is impossible. Thus, we have a situation with *partial interference*.

These methods require that there are multiple groups. They also require that the groups are fixed, i.e., that individuals belong to one and only one group during the study period.

There are methods for randomized studies and in the last few years methods for non-randomized studies have also been suggested.

3.1.1 Randomized studies

Hudgens and Halloran (2008) is one of the first and also one of the most cited papers suggesting ways to make inference under interference in randomized studies. Many of the subsequent proposals are continuations on this seminal paper, which we, therefore, will describe in quite some detail.

Estimating direct, indirect, total, and overall causal effects in two-stage randomized experiment

Hudgens and Halloran (2008) defined causal estimands which they denote direct, indirect, total, and overall causal effects in a two-stage randomized setting. The population of interest consists of $N > 1$ groups where interference is possible between individuals in the same group but not between individuals in different groups. The vector $\mathbf{Z}_i \equiv (Z_{i1}, \ldots, Z_{in_i})$ contains information on the treatment assignment (0 or 1) of each of the n_i individuals in group i, $i = 1, \cdots N$. The potential outcome of individual j in group i is denoted by $Y_{ij}(\mathbf{z}_i)$, where $\mathbf{z}_i$ are possible values of $\mathbf{Z}_i$. Thus, the outcome of individual j depends on the treatment assignment of individual j, Z_{ij}, as well as on the other $n_i - 1$ individual's assignments, $\mathbf{Z}_{i(j)}$, but the outcome of individual j does not depend on the treatment assigments of the individuals in other groups.

In the first stage of the randomization, the N groups are randomized to either strategy ψ with a high proportion treated or strategy ϕ with a low proportion treated individuals (or even no treated individuals[4]). In the second stage, individuals within groups are randomly assigned to be treated ($z = 1$) or not treated ($z = 0$), where individual treatment probabilities depend only on whether the group is assigned to strategy ψ or ϕ in the first stage of randomization. That is, within a group all individuals have

[4]If ϕ is a strategy implying no treated individuals then a group randomized to ϕ is called a *no-intervention* group.

the same treatment probability but that probability is different between the two group strategies.

Direct causal effects

The individual direct causal effect for individual j in group i is defined as the difference in potential outcomes for that individual given the treatment and without the treatment, all else being equal. That is,

$$CE_{ij}^{D}(\mathbf{z}_{i(j)}) \equiv Y_{ij}(\mathbf{z}_{i(j)}, z_{ij} = 1) - Y_{ij}(\mathbf{z}_{i(j)}, z_{ij} = 0).$$

The individual average direct causal effect for individual j in group i under strategy ψ is defined as the difference in average potential outcomes of treatment compared with no treatment, where the average is over all possible treatment allocations of the other $(n_i - 1)$ individuals. We denote this

$$\overline{CE}_{ij}^{D}(\psi) \equiv \overline{Y}_{ij}(1;\psi) - \overline{Y}_{ij}(0;\psi).$$

Averaging over all individuals of the group gives the group average direct causal effect:

$$\overline{CE}_{i}^{D}(\psi) \equiv \overline{Y}_{i}(1;\psi) - \overline{Y}_{i}(0;\psi) = \sum_{j=1}^{n_i} \overline{CE}_{ij}^{D}(\psi)/n_i.$$

The population average direct causal effect is defined as the average of the group average direct causal effect over all groups:

$$\overline{CE}^{D}(\psi) \equiv \overline{Y}(1;\psi) - \overline{Y}(0;\psi) = \sum_{i=1}^{N} \overline{CE}_{i}^{D}(\psi)/N.$$

The two first direct causal effects (individual and individual average) are not estimable as an individual is only observed under either treatment ($z_{ij} = 1$) or no treatment ($z_{ij} = 0$). The estimator of the group average direct effect is just the difference of sample means between the treated an non-treated individuals of group j. Taking averages of this estimate over all groups under strategy ψ yields an estimate of the population average direct causal effect.

The effects are defined and estimated in the same manner under strategy ϕ.

Indirect causal effects

An indirect causal effect on an individual is an effect from the treatment received by others in the same group, i.e., a spillover effect. The individual indirect causal effect is defined as the difference between the outcome an untreated individual would have in a group under strategy ψ and the outcome the same individual would have in a group under strategy ϕ. We write this as

$$CE^I_{ij}(\mathbf{z}_{i(j)}, \mathbf{z}'_{i(j)}) \equiv Y_{ij}(\mathbf{z}_{i(j)}, z_{ij} = 0) - Y_{ij}(\mathbf{z}'_{i(j)}, z'_{ij} = 0),$$

where $\mathbf{z}_i$ ($\mathbf{z}'_i$) is the treatment indicator vector of the ψ (ϕ) group. Similarly to direct effects, individual average, group average, and population average indirect effects are defined as

$$\overline{CE}^I_{ij}(\psi, \phi) \equiv \overline{Y}_{ij}(0, \psi) - \overline{Y}_{ij}(0, \phi),$$

$$\overline{CE}^I_{i}(\psi, \phi) \equiv \overline{Y}_{i}(0, \psi) - \overline{Y}_{i}(0, \phi) = \sum_{j=1}^{n_i} \overline{CE}^I_{ij}(\psi, \phi)/n_i$$

and

$$\overline{CE}^I(\psi, \phi) \equiv \overline{Y}(0, \psi) - \overline{Y}(0, \phi) = \sum_{i=1}^{N} \overline{CE}^I_{i}(\psi, \phi)/N,$$

respectively.

The individual indirect causal effects are not possible to estimate due to the fact that an individual belongs to either a group under ψ or a group under ϕ. Also at the group level, the indirect causal effect cannot be estimated due to the fact that a group is only observed under one of the two strategies. The population average indirect causal effect is estimated by the difference in average sample means of the untreated in the ψ groups and the untreated in the ϕ groups.

Total causal effects

A total causal effect is the difference in outcomes between being treated in a group under strategy ψ and being untreated in a group under strategy ϕ. The individual, individual average, group average, and population average total causal effects are defined by

$$CE^T_{ij}(\mathbf{z}_{i(j)}, \mathbf{z}'_{i(j)}) \equiv Y_{ij}(\mathbf{z}_{i(j)}, z_{ij} = 1) - Y_{ij}(\mathbf{z}'_{i(j)}, z'_{ij} = 0),$$

$$\overline{CE}_{ij}^{T}(\psi,\phi) \equiv \overline{Y}_{ij}(1,\psi) - \overline{Y}_{ij}(0,\phi),$$

$$\overline{CE}_{i}^{T}(\psi,\phi) \equiv \overline{Y}_{i}(1,\psi) - \overline{Y}_{i}(0,\phi) = \sum_{j=1}^{n_i} \overline{CE}_{ij}^{T}(\psi,\phi)/n_i$$

and

$$\overline{CE}^{T}(\psi,\phi) \equiv \overline{Y}(1,\psi) - \overline{Y}(0,\phi) = \sum_{i=1}^{N} \overline{CE}_{i}^{T}(\psi,\phi)/N,$$

respectively. Note that the total effect is the sum of the direct and indirect effects on each level (i.e., individual, individual average, group average, and population average levels).

Again, the individual and group-level effects cannot be estimated as individuals/groups are only observed under one of the two treatments/strategies. The population average total causal effect can be estimated by taking averages of the sample means of the treated in the ψ groups to estimate $\overline{Y}(1,\psi)$ and by taking averages of the sample means of the untreated in the ϕ groups to estimate $\overline{Y}(0,\phi)$.

Overall causal effects

The overall causal effect is the effect of strategy ψ compared to strategy ϕ on the outcomes on individual, individual average, group average and population average level defined by

$$CE_{ij}^{O}(\mathbf{z}_i,\mathbf{z}_i') \equiv Y_{ij}(\mathbf{z}_i) - Y_{ij}(\mathbf{z}_i'),$$

$$\overline{CE}_{ij}^{O}(\psi,\phi) \equiv \overline{Y}_{ij}(\psi) - \overline{Y}_{ij}(\phi),$$

$$\overline{CE}_{i}^{O}(\psi,\phi) \equiv \overline{Y}_{i}(\psi) - \overline{Y}_{i}(\phi)$$

and

$$\overline{CE}^{O}(\psi,\phi) \equiv \overline{Y}(\psi) - \overline{Y}(\phi).$$

In practice, only the population average total causal effect can be estimated. This is done by averaging the group means of the ψ and ϕ groups and taking the difference between those averages.

The overall causal effect can often be of interest for policy makers. Examples of research questions that could be answered include 'how would

infection rates, on average, differ between two vaccination schemes?' and 'how would the unemployment rate change if a new training program is implemented?'.

Confidence intervals for the direct, indirect, total, and overall causal effects

Under the additional assumption of *stratified interference*, Hudgens and Halloran (2008) suggested estimators of the variance of the direct, indirect, total, and overall causal effect estimators too. In short, the assumption of stratified interference can be seen as an assumption that only the proportion of treated in the group matters for the potential outcomes of an individual and not which particular individuals in the group that are treated.

For binary outcomes, exact confidence intervals for the four effects defined in Hudgens and Halloran (2008) were derived in Tchetgen Tchetgen and VanderWeele (2012). However, Rigdon and Hudgens (2015) pointed out that these intervals are rather conservative and can be very wide. Instead these authors derived other exact confidence intervals that perform better.

Liu and Hudgens (2014) considered large sample confidence intervals of Wald-type. These perform well when the number of clusters are large, which, however, is seldom the case in practice.

The `R` package `interferenceCI` (Rigdon, 2015) can be used to compute all the confidence intervals mentioned above.

3.1.2 Non-randomized studies

For clustered data from non-randomized studies there are also a few suggestions, e.g., Tchetgen Tchetgen and VanderWeele (2012), Lundin and Karlsson (2014), and Ferracci et al. (2014). Partial interference is assumed, in the same manner as in Hudgens and Halloran (2008). All three papers also make additional assumptions similar to those usually made for causal inference in observational studies, e.g., assuming that all confounders are controlled for (unconfoundedness assumptions). In the estimations of causal effects, estimated treatment probabilities (propensity scores), i.e., the conditional probability for an individual to be treated given his/her charachteristics, are used to make individuals comparable either by matching methods or by using propensity scores as weights in inverse probability weighting (IPW) estimators (e.g., Hirano et al., 2003).

Other, more parametric, methods on how to handle interference in non-randomized studies with clustered data are found in Hong and Raudenbush

(2006) and Verbitsky-Savitz and Raudenbush (2012), which both used parametric generalized hierarchical linear models to mimic multi-stage randomized experiments. These methods are not covered in this review.

Two-stage partly non-randomized studies

Lundin and Karlsson (2014) considered two-stage experiments where the assignment of treatment assignment strategies (ψ, ϕ) to clusters is randomized (first stage) but the assignment of individuals within clusters to treatment or control (second stage) is not. Instead they are assigned in some other manner, e.g., by self-selection or competition to the fixed number of available positions for treatment, determined by the randomly selected treatment assignment strategy for their cluster.

Under additional assumptions of unconfoundedness and overlap in the covariate distributions of treated and untreated individuals within each group, the authors proposed that individual treatment probabilities should be estimated using background information on the individuals and that these estimated probabilities then should be used to construct IPW estimators of average potential outcomes needed to estimate the direct, indirect, total, and overall causal effects defined in Hudgens and Halloran (2008).

The method proposed in Lundin and Karlsson (2014) can, for example, be used when evaluating a labor training program in which some geographical regions have been randomly selected to administer a program with a fixed number of training positions, but the participating unemployed within these regions are selected in a non-random way.

Lundin and Karlsson (2014) illustrated the method by evaluating the effect on childrens behaviour of implementing a parenting support program, Triple P, in some preschools in Uppsala, Sweden. The preschools that participated in the study were randomized to be "intervention preschools", i.e., to make available the Triple P to the parents of children within the preschool, or to be "control preschools", i.e., not offering the Triple P to the parents. Within the intervention preschools parents could themselves choose to participate in the program or not.

Two-stage non-randomized studies with a continuous treatment at the cluster level

In Ferracci et al. (2014) the proportion treated in a market (market being the word used instead of cluster or group there) is considered as the factor that besides the individuals own treatment statuses affects the outcomes of

individuals within that market. The potential outcomes of individual j in group i can then be written as $Y_{ij}(z_{ij}, q_i)$, i.e., as a function of his/her own treatment assignment ($z_{ij} = 0$ or $z_{ij} = 1$) and the proportion treated in the group, q_i ($0 \leq q_i \leq 1$).

Their proposal can also, similar to Lundin and Karlsson (2014), be seen as an extension of the proposal in Hudgens and Halloran (2008). However, the group level treatment assignment strategies, i.e, the proportion treated in a group, is no longer limited to a fixed number (ψ or ϕ) as in Hudgens and Halloran (2008) (and Lundin and Karlsson, 2014) but considered as a continuous "group level treatment" variable. Also, neither the proportion treated in a group nor which individuals within a group that receive treatment are decided by randomization.

Still, the aim in Ferracci et al. (2014) is to estimate population average potential outcomes, $\overline{Y}(1, q)$ and $\overline{Y}(0, q)$ for all q. Then, by taking different contrasts between these estimated population average potential outcomes, population average direct causal effects (differences between estimated average potential outcome under treatment and estimated average potential outcome under control for the same values of q) and population average indirect causal effects (differences between estimated average potential outcomes under control for different values of q) can be calculated.

The population average potential outcomes are suggested to be estimated with a two-step procedure[5], which is described in great detail and illustrated with a labour market evaluation study in Ferracci et al. (2014). In short, they suggest to first consider the proportion treated in each group as fixed, and to use weighted regression with estimated propensity scores as weights to estimate the group average potential outcomes within each group separately. Then, considering the groups as the units of observation, the group average potential outcomes as the units' potential outcomes, and the proportion treated as the continuous treatment variable, they suggest using a flexible regression of the group average potential outcome on the proportion treated and an estimated *generalized propensity score* (Hirano and Imbens, 2004) to get estimates of the population average potential outcomes.

[5]The required identifying assumptions are also formed in two steps. The identifying assumptions basically say that: i) there exists, within each group, observable individual covariates that, if conditioned upon, make the treatment assignment independent of the potential (individual) outcomes and ii) there exists observable group covariates that, if conditioned upon, make the "group level treatment", i.e., the assignment of proportion treated in the groups, independent of the group average potential outcome.

"One-stage" non-randomized studies

Tchetgen Tchetgen and VanderWeele (2012) also extended the estimators of Hudgens and Halloran (2008) by proposing IPW estimators in observational studies. However, they do not assume that treatment allocation is made in two steps. Instead they assume that each individual's probability of being treated is a function of a covariate vector, which includes both individual level and cluster level covariates.

To estimate the direct, indirect, total, and overall causal effects using an IPW estimator, Tchetgen Tchetgen and VanderWeele (2012) assume that conditional on the covariates, the treatment allocation is independent of the potential outcomes and that all possible treatment allocations for a group has a strictly positive probability of occurring. Thus, Tchetgen Tchetgen and VanderWeele (2012) make group-level versions of the unconfoundedness and overlap assumptions.

Note that they do not consider the treatment assignment strategy as a continuous variable as in Ferracci et al. (2014); at least they do not model and estimate the causal effects as (smooth) functions of the proportion treated.

3.2 Network data

In many practical situations it is not possible to consider individuals as belonging to separated groups as in the previous section, e.g., when all individuals belong to a single cluster or if the clusters are not separated enough. In these situations, there is risk for "global interference" instead of partial interference.

If the relations or social connections between individuals are known, we can consider each individual as a vertex in a graph or network and each connection as an edge in the same network. In this section, we review some of the proposals made for causal inference in such social networks but first we describe the process of experiments in networks.

The description of the process of experiments in networks closely follow the description in Eckles et al. (2014), which we think is very explicatory. Let $G = (V, E)$, where V is the vertex set and E is the edge set, denote the network.

The process of experiments in networks consists of four phases:

1) initialization,

2) treatment assignment,

3) outcome generation, and

4) analysis.

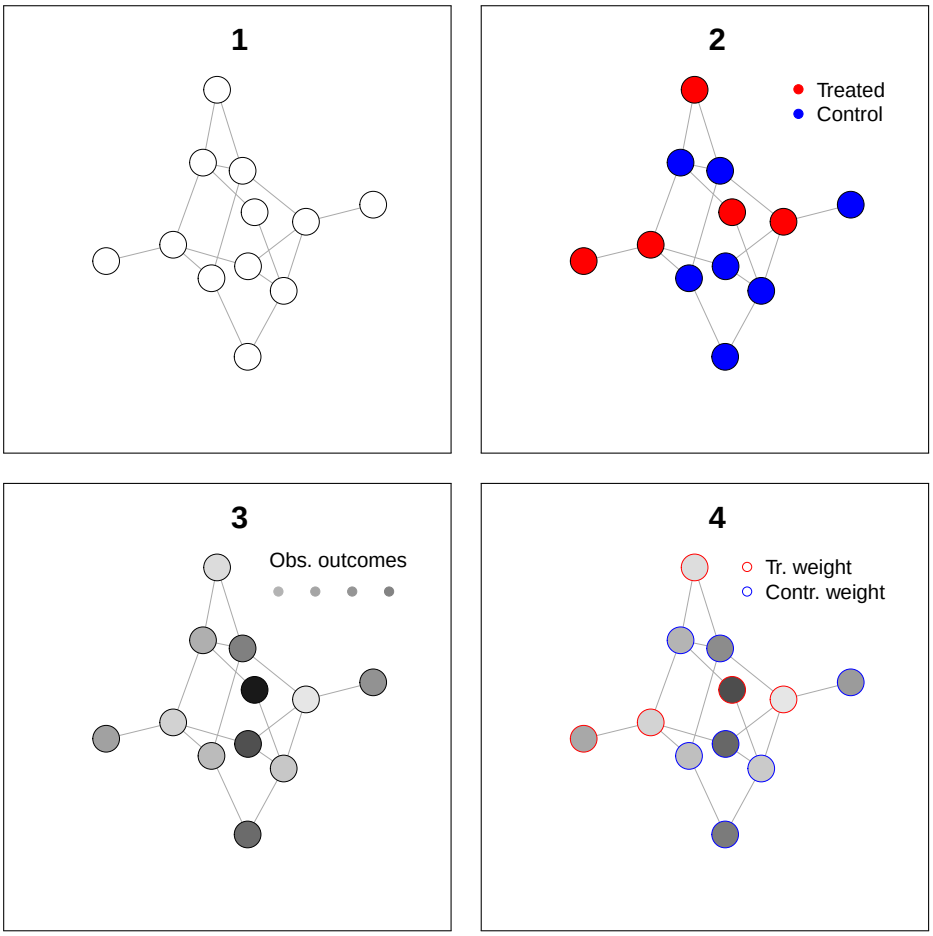

Figure 1: The process of experiments in networks consists of 1) initialization, 2) treatment assignment, 3) outcome generation, and 4) analysis. *Adapted from Figure 1 in Eckles et al. (2014).*

The *initialization phase* consists of defining the graph and collecting information about the vertices, i.e., the individuals' covariates and behaviour prior to the experiment. After the initialization phase we have a particular

network $G = (V, E)$ and, possibly, a collection of vertex characteristics, $\mathbf{X}$. In the *treatment assignment phase* treatments are assigned to each vertex according to some experimental design. Given the network and treatment assignment, some data generating process produces the observed outcomes in the *outcome generation phase*. Finally, in the *analysis phase* an estimator is constructed, assigning different weights to the observations, in order to estimate the effects of interest, i.e., the estimands. The process of experiments in networks is illustrated in Figure 1.

3.2.1 Estimating the average treatment effect of global treatment vs global control

When considering if an intervention would be beneficial if applied to all individuals, a natural choice would be to look at the average treatment effect of "applying the treatment to all units [in the network] compared with applying the control treatment to all units" (Eckles et al., 2014, p. 2). We henceforth call this estimand the average treatment effect of global treatment (ATEGT).

Eckles et al. (2014) focused on how experimental design choices and/or analysis decisions might reduce the bias of the estimator of ATEGT. They provide sufficient conditions for bias reduction by using *graph cluster randomization* (experimental design) and by using estimators that define *effective treatments* (analysis method) under potential global interference.

In short, graph cluster randomization means that the vertices are partitioned into clusters. Then, by a Bernoulli trial, each cluster is assigned to be either a "treatment cluster" or a "control cluster". All the vertices in the "treatment clusters" are treated while all the vertices in the "control clusters" are given the control treatment. Hence, when a vertex is assigned treatment (or control) by graph cluster randomization so are also the vertices close to it in the network, by design. The intuition is that graph cluster randomization results in a situation closer to the situation of interest (i.e., all vertices under treatment or all vertices under control) than, e.g., the independent random assignment of units ($Z_j \sim \text{Bernoulli}(p)$), which is the most commonly used assignment design (cf. Figure 2).

Using estimation methods that define effective treatments are, in Eckles et al. (2014), described as using only data from vertices that are effectively in global treatment or effectively in global control to estimate ATEGT. For example, an estimator of ATEGT might only compare vertices in treatment that are surrounded by vertices in treatment with vertices in control that are surrounded by vertices in control. Again, the intuition is that this results in

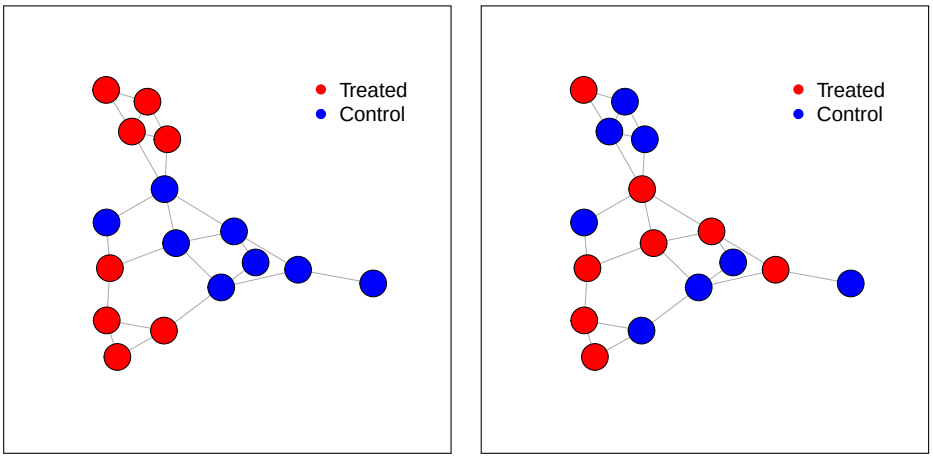

Figure 2: Examples of treatment assignment designs: graph cluster randomization (left panel) and independent random assignment (right panel).

a situation closer to the situation of interest.

Eckles et al. (2014) investigated the performance of the aforementioned methods by means of simulation for different networks and different outcome generating processes, which are based on treatment assignments, vertex covariates, and neighboring vertices' covariates. The simulation results indicate that using graph cluster randomization can reduce the bias substantially compared to independent random assignment. Moreover, simulation results indicate that using specific estimators can reduce bias even if they are based on incorrect definitions of effective treatments. Reduction in bias usually comes at the cost of increased variance but in many of the situations in their paper the bias reduction is large enough to reduce root mean squared error (RMSE), which is a function of both bias and variance of the estimator.

It is worth mentioning that the authors' purpose with this paper, in contrary with many others papers by other authors, was to evaluate the methods under realistic conditions rather than deriving those (maybe unrealistic) conditions that would make the methods unbiased.

3.2.2 Estimating effects of peer's treatment assignment by exposure mapping

In Aronow and Samii (2015) it is suggested that instead of trying to estimate causal effects of treatment one could estimate causal effects of *treatment exposure*. Individuals have a potential outcome per each possible treatment exposure (level) instead of one per each treatment, which is what is assumed under the no-interference assumption. The causal effects of interest could, for example, be the difference between average outcomes under two different treatment exposure levels.

Treatment exposure levels are defined (mapped) from the treatment assignment and from the constitution of the network at hand. Every vertex is potentially exposed by its own assigned treatment and by the assigned treatments to the other vertices in the network. Which of the other vertices' treatments that do affect its outcome depends on the social structure described by the network and how the particular treatment effect can transmit through it. The latter is an issue for the researcher to decide about; some judgement has to be made about how peripheral social connections between individuals have to be in order to not contribute with treatment exposure to each other. Perhaps it is plausible to assume that only a vertex's closest neighbors may effect it and not its neighbors' neighbors for example. Perhaps the exposure level can be decided based on the number of treated neighbors without having to consider which of the neighbors that are treated, i.e., that all neighboring vertices are on equal footing regarding transmitting treatment exposure.

Not just looking at treatments assigned by the randomization, but also at treatment exposures received, means that the process of experiments in networks consists of an additional phase compared to the process illustrated in Figure 1. In the *exposure mapping phase*, indirect exposure of the treatments, which have been randomly assigned in the treatment assighment phase, are transmitted through the network. Given the exposure level, some data generating process produces the observed outcomes in the outcome generation phase. Thus, the exposure mapping phase takes place between phase 2 and 3 in Figure 1.

In Figure 3, where the process of experiments with exposure mapping in networks is illustrated, we let the exposure mapping phase be denoted 2½ to emphasize that it takes place between phase 2 and 3 in Figure 1.

In order to illustrate this we consider an example, where indirect exposure transmits to a vertex's neighbors and the amount of indirect exposure is the same regardless of how many of a vertexs neighbors that are assigned

to treatment. This means that each vertex falls into exactly one of four exposure conditions (levels):

I *Direct and indirect exposure:* the vertex has been assigned to treatment and at least one of its neighbors has also been assigned to treatment.

II *Isolated direct exposure:* the vertex has been assigned to treatment but all its neighbors have been assigned to control.

III *Indirect exposure:* the vertex has been assigned to control but at least one of its neighbors has been assigned to treatment.

IV *No exposure:* neither the vertex nor its neighbors have been assigned to treatment.

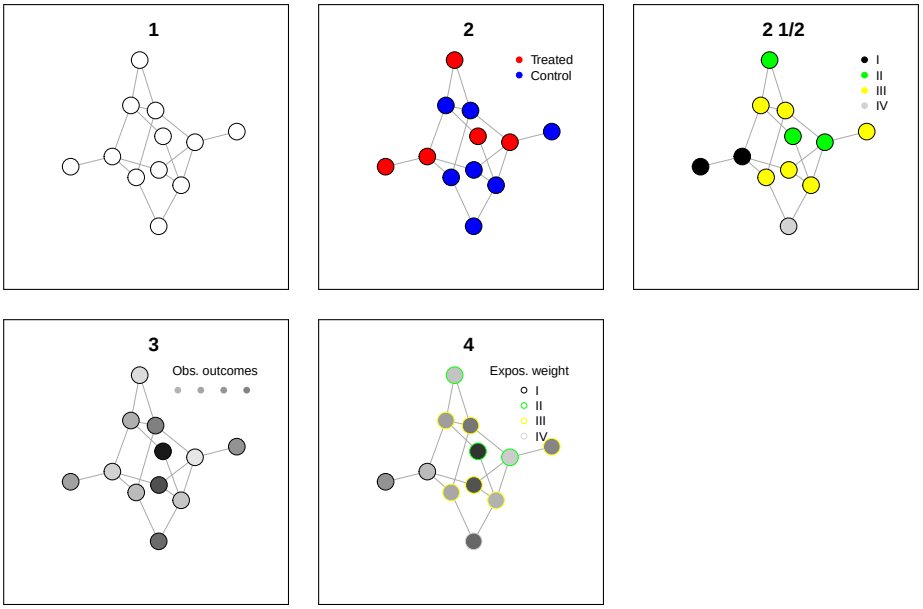

Figure 3: The process of experiments with exposure mapping in networks consists of 1) initialization, 2) treatment assignment, 2½) exposure mapping, 3) outcome generation, and 4) analysis. In the exposure mapping phase treatment exposure transmits through the network, which (for this example) yields four exposure levels: I) Direct and indirect exposure, II) isolated direct exposure, III) indirect exposure, and IV) no exposure.

The example above is similar to the simulation study set up in Aronow and Samii (2015), but they also describes other situations that yields different exposure levels. In Hellman and Lindberg (2015) [bachelors thesis] yet other examples are given.

When the network, the randomization design, and the principles of the exposure mapping are known exactly, the probabilities of being exposed to one or another level are also, for each vertex, known. For the example above, the probabilities that vertex j belongs to exposure class I, II, III, and IV respectively are known. These are called the *individual exposure probabilities*. Also the *joint exposure probabilities*, i.e., probabilities that the pair of vertices j and j' belongs to specific exposure levels, are known. If the network is too large to be able to compute the exposure probabilities exactly, it is possible to approximate them by simulation; taking the relative frequencies of the vertex belonging to an exposure class in repeated assignments of treatments according to the randomization design.

In the example, each individual j has four potential outcomes and we want to estimate the average potential outcome in the population for each of these four exposure levels and then look at interesting contrasts among them. However, we can not calculate these averages directly, since we only observe the potential outcome of the exposure level that the individual actually is exposed to and not the other three outcomes. Suppose all units have non-zero probabilities of being exposed by exposure level, l, for all l (l =I, II, III, IV in the example). Then, by design[6], the set of individuals for which we observe the potential outcome of exposure level l is an unequal-probability without-replacement sample from the set of all individuals potential outcomes of exposure level l. Thus, a Horvitz-Thomson (HT) estimator[7] (Horvitz and Thompson, 1952) can be used to estimate the average potential outcome.

However, as Aronow and Samii (2015) points out, even if treatment assignment can be manipulated arbitrarily with the experimental design, treatment exposure levels may be constrained by the characteristics of the network, i.e., the social relations amongst individuals. For example, in the example above, a vertex without neighbors can never be exposed to level I and III and a vertex with many neighbors may have very low probability of being exposed to level II or IV. In these cases, the method will run into problems.

[6] Phases 1 and 2 in the process of experiments in networks, see Figure 1.

[7] In causal inference the term IPW estimator is a more commonly used term than the HT estimator (cf. Hirano et al., 2003).

3.2.3 Estimating k-level peer influence effects

Toulis and Kao (2013) also considers experiments in networks and introduce new estimands that they think are of interest when considering peer influence effects. These estimands are based on the number of neighbors, k, that an individual (vertex) has. In order to describe these estimands, we first introduce some notation from their paper.

For vertex $j \in V$, let $\mathcal{N}_j$ denote its neighborhood (with the vertex j itself excluded), i.e., the set of all vertices that are neighbors to j (have an edge between them and vertex j). Let $\mathbf{Z}_{\mathcal{N}_j}$ be the treatment assignment vector of the neighbors of j. Assume that, for all j, $Y_j(\mathbf{Z}) = Y_j(Z_j, \mathbf{Z}_{\mathcal{N}_j})$, i.e., that a vertex outcome can be affected only by the treatment it receives itself and by the treatments its neighbors receive.

Let V_k be the set of vertices that have at least k neighbors. Also, let $\mathcal{M}_{jk}$ be the set of neighbors of $j \in V_k$, who are also neighbors to at least one other vertex in V_k. We call $\mathcal{M}_{jk}$ for *shared neighbors* of j. Neighbors not shared with others is called *insulated* neighbors.

Let $\mathbf{Z}(\mathcal{N}_j; k)$ denote the set of all assignments on $\mathcal{N}_j$ with exactly k neighbors of j being treated. Vertex j is then said to be *k-level exposed*. There are $\binom{|\mathcal{N}_i|}{k}$ such possible assignments, where $|\mathcal{N}_i|$ denotes the cardinality (i.e., the number of elements) of the set $\mathcal{N}_i$. Let $\mathbf{Z}_1(\mathcal{N}_j; k)$ denote the set of all assignments in $\mathbf{Z}(\mathcal{N}_j; k)$, where at least one of the shared neighbors of j is treated, while $\mathbf{Z}_0(\mathcal{N}_j; k)$ denotes the set of all assignments in $\mathbf{Z}(\mathcal{N}_j; k)$, where j is k-level exposed but all of its shared neighbors receives the control treatment, i.e., $\mathbf{Z}_0(\mathcal{N}_j; k) = \mathbf{Z}(\mathcal{N}_j; k) \setminus \mathbf{Z}_1(\mathcal{N}_j; k)$.

The first estimand introduced in Toulis and Kao (2013) is the estimand for *primary effects*

$$\xi \equiv \frac{1}{|V|} \sum_j Y_j(1, \mathbf{Z}_{\mathcal{N}_j} = \mathbf{0}) - Y_j(0, \mathbf{Z}_{\mathcal{N}_j} = \mathbf{0}),$$

i.e., the primary effect is the average effect of treatment vs control when all other vertices in the neighborhood are assigned to control. Thus, the average treatment effect when there is no peer influence at all, since no peers are treated.

They also introduces estimands for peer influence effects, both main estimands (δ_k) and additional estimands ($\delta_{k,1}, \delta_{k,0}$). The (main) estimand for *k-level peer influence effects* is defined as

$$\delta_k \equiv \frac{1}{|V_k|} \sum_{j \in V_k} \left[\binom{|\mathcal{N}_j|}{k}^{-1} \sum_{\mathbf{z} \in \mathbf{Z}(\mathcal{N}_j; k)} Y_j(0, \mathbf{z}) - Y_j(0, \mathbf{0}) \right],$$

i.e., for the untreated, the effect of being k-level exposed compared to not being exposed to peer influence at all (all neighbors being untreated).

The (additional) estimands for *k-level peer influence effects of insulated neighbors* are defined as

$$\begin{aligned}\delta_{k,0} &\equiv \frac{1}{|V|}\sum_j \binom{|\mathcal{N}_j| - |\mathcal{M}_{jk}|}{k}^{-1} \\ &\times \sum_{\mathbf{z}\in\mathbf{Z}_0(\mathcal{N}_j;k)} Y_j(0,\mathbf{z}) - Y_j(0,\mathbf{0}),\end{aligned} \tag{1}$$

and for *k-level peer influence effects of non-insulated neighbors* as

$$\begin{aligned}\delta_{k,1} &\equiv \frac{1}{|V|}\sum_j \left(\binom{|\mathcal{N}_j|}{k} - \binom{|\mathcal{N}_j| - |\mathcal{M}_{jk}|}{k}\right)^{-1} \\ &\times \sum_{\mathbf{z}\in\mathbf{Z}_1(\mathcal{N}_j;k)} Y_j(0,\mathbf{z}) - Y_j(0,\mathbf{0}).\end{aligned} \tag{2}$$

Thus, in (1), for the untreated, the effect of being k-level exposed by its insulated neighbors (neighbors not shared with others) compared to not being exposed to peer influence at all. In (2), also for the untreated, the effect of being k-level exposed by non-insulated neighbors (at least one of the treated neighbors is a shared neighbor) compared to not being exposed to peer influence at all.

Toulis and Kao (2013) also propose ways to estimate their suggested estimands. Their proposals are rather complicated and therefore, a detailed description of them is beyond the scope of this paper. Their first proposal is a randomization based inference approach and they argue that a *sequential randomization design* should be used when conducting the experiment in order to get enough vertices in the "status of interest", e.g., non-exposed or k-level exposed (by insulated or non-insulated neigbours) and so on. This requires that the network is known but also it depends heavily on the network topology, i.e., the how the social relations amongst individuals are composed. For example, consider, as an extreme case, a network where all vertices are connected with all the other vertices, then as soon as a vertex get k-level exposed no other vertex can be non-exposed. Their second proposal is a model-based approach, viz. a linear model which relies on an assumption of additivity of the primary effects and peer influence effects to the response mean. In the model, they allow for network uncertainty by considering weighted random networks. We refer our readers to their paper for further details.

4 Empirical examples of interference

In this section we portray a number of evaluation studies where the investigators have found indirect effects. The examples include both labor market program evaluations and evaluations of other interventions, e.g., in schools, neighborhoods or workplaces. The purpose of this section is to show that the problem of interference is present in many situations and that it should not be neglected in evaluation studies.

4.1 Displacement effects from labor market policies in France

Crépon et al. (2013) found that the significant and positive treatment effect of "job placement assistance" for young, educated job seekers on finding a stable job, that was seen (assuming no-interference) in a two-step randomized experiment in France was actually not as good as it first seemed. The job seekers that were randomly selected to get job placement assistance were indeed employed to a greater extent than the ones without the treatment, but it was not only due to that the individuals under treatment got a bigger chance to get employed but rather that the untreated individuals got a smaller chance when other job seekers in their region got treatment.

In the experiment, 235 regions participated, and the treatment was distributed to the job seekers through a two-step procedure. The regions were randomly assigned to offer 0%,25%, 50% or 100% of the job seekers in the region the job placement assistance. Then, based on the result of the first step of the experiment, a proportion of job seekers in each region were randomly selected to be offered the job placement assistance.

Crépon et al. (2013) reported that, after eight months, the job seekers assigned to job placement assistance were 2.5 percentage points more likely to have a stable job than unassigned job seekers in treatment areas. Hence, the program seemed to have a positive effect. At the same time, untreated job seekers in a treated area were 2.1 percentage points less likely to find any stable job than untreated job seekers in control areas, i.e., areas with 0% treated individuals. This so called displacement effect[8] is reported to be significant at the 10% level. Thus, the job placement assistance program had little net benefits. However, they could not find enough empirical evidence to say that the effect on untreated job seekers was different in areas with different proportion of treated individuals; the displacement effect was about the same size in areas with 25% treated and areas with 75% treated.

[8] Or indirect effect, if using the terminology introduced in Section 3.1.1 .

Ferracci et al. (2014) also analyzed data from regions in France. They considered "participation in a training program" as the treatment and used register data on the proportions treated in different markets (defined by region, occupation, and time period) over several years. They assumed that there is no spillover between markets and that the individuals potential outcomes can be written as a function of its own treatment status and the proportion of treated individuals in the market it belongs to. The method used to analyze the data is the method suggested in the same paper (cf. Section 3.1.2).

The authors argue that SUTVA violations are to be expected in this situation, since if many individuals are treated there may be crowding out among workers but also that if many individuals are treated (i.e., that the number of qualified job seekers increases) there may be more vacancies posted by the firms. Indeed, Ferracci et al. (2014) report that the estimated average potential outcomes do not remain constant when the proportion of treated changes. The estimated average probability of getting a job within a year if treated, decreases when the proportion of treated increases. The estimated average probability of getting av job within a year if untreated, first decreases when the proportion of treated increases, but then starts increasing when the proportion treated is high enough (5.5%), i.e., it follows a convex pattern.

4.2 Indirect effects of foreign ownership on firm productivity in China

Girma et al. (2015) analyzed firm level data from the Chinese manufacturing industry, and the effect of having foreign ownership instead of being a domestic firm on firm productivity.

The firms were divided into 127 clusters, based on geographic areas and industry classification, and it was assumed that the no-interference assumption holds across clusters, but not within cluster.

The potential outcomes under the two treatments (foreign or domestic ownership) were expressed as functions of the individual firms treatment status and the proportion of treated firms in the cluster that the firm belonged to. For details about the estimation procedure, which involves measures taken to deal with the treatment not being randomized (e.g., estimating the propensity of a firm being foreign owned), see Girma et al. (2015).

The results show that the potential outcomes vary systematically with the proportion of treated in the cluster. The estimated direct effect was positive and higher the more foreign-owned firms there were in a cluster.

The estimated indirect effects of foreign-owned firms in the cluster on the domestic firms were negative and differed with the proportion of foreign ownership in the cluster. Spillover was more negative with increasing proportion of foreign ownership up to a threshold when spillover became less negative. The authors argue that their work provides important inputs into the policy debate on the benefits from agglomerations of foreign owned firms.

4.3 Work absence influenced by peers behaviour in Sweden

In Sweden, workers receive temporary benefits for sickness absence from the workplace up to seven days without being required to present a doctor's certificate. In a randomized field experiment, conducted in 1998 in Gothenburg, Sweden, individuals assigned to treatment were allowed to be absent up to fourteen days without presenting a doctor's certificate while the individuals assigned to the control still had to show a doctors certificate on the 8th day of a sickness absence spell in order to continue to receive temporary benefits. The effect of the revised rules on the length of the non-monitored sickness absence spells was of interest to study.

Johansson et al. (2014) analyzed the results of the experiment and found that the decreased monitoring of absenteeism increased non-monitored absence among the treated workers. They also found a positive peer effect in absenteeism, i.e., the higher the proportion treated individuals in a workplace, the more non-monitored absence increased among untreated individuals. Hesselius et al. (2009) and Hesselius et al. (2013) also analysed the same data and draw similar conclusions about the existence of peer effects in the experiment.

4.4 Peer influence in an anti-conflict intervention in schools in New Jersey, USA

Paluck et al. (2016) report results from an experiment in 56 schools in New Jersey, USA, investigating the effect of an anti-conflict (anti-bullying) intervention performed by individual students on the whole schools' behavioral climate, measured with social norm questions regarding conflict behaviors asked before and at the end of the experiment and with schools' administrative records of student conflict-related disciplinary events. At the beginning of the study, each school's social network were mapped by asking all students to report up to 10 students at their school whom they choosed to spend time with (in school, out or school, or online) during the last few weeks.

Half of the schools were randomized to receive an anti-conflict intervention (henceforth called treatment schools) and there rest were assigned to be control schools. At each treatment school, a group of students (so called "seed-eligibles") were selected by a deterministic algorithm based on gender and grades (see details in Paluck et al., 2016). By randomization, half of the seed-eligibles were invited to participate in the anti-conflict intervention, i.e., to be "seeds". If a seed was in the top 10% of their school when it came to number of connections to other students (measured as number of reported connections by other students) it was called a "social referent seed". The proportion of social referent seeds varied between treatment schools, from 0-37%. During the anti-conflict intervention, the seeds were encouraged by a trained research assistant (with whom they met every other week) to take public stance against different types of conflicts, which they had identified at their own school. For example, seed students created hashtag slogans and posted them online together with their own photos to link the statements to their identities.

Due to the randomization of schools and seeds (from the group of seed eligibles) the evaluation of school-level outcomes were straight forward using linear regression, see details in Paluck et al. (2016). The average levels of disciplinary reports of student conflicts in treatment schools were significantly lower than in control schools. In control schools every student was disciplined for student conflicts on average 0.2 times during the year and in treatment schools on average 0.14 times. Schools with the highest proportion of social referents seeds assigned to treatment had the greatest decline in number of disciplinary reports, e.g., with 20% of the seeds in treatment schools being social referent seeds each student was disciplined for student conflicts on average 0.08 times during the year.

Paluck et al. (2016) also considered how seed students (i.e., the students who participated in the anti-conflict intervention by taking public stance against conflicts) affected other students in their social network in terms of how they answered the social norm questions about conflict behavior at the end of the experiment. Since the network was known before the randomization of seeds and schools it was possible to calculate, for each student in a school network, the probability of being exposed to a social referent seed, to a non-referent seed, or to no seed at all. The authors consider four exposure levels (see details in Paluck et al., 2016) and restrict their analysis on the sub-population of students which had a positive probability of falling into all four exposure levels. They use a IPW estimator to estimate the average potential outcome under these four different exposure levels, cf. Aronow and Samii (2015), and found "a significant social effect attributable to seed

students, and particularly as a result of social referent" (Paluck et al., 2016, p. 569).

The authors therefore conclude that their study shows that it is possible to reduce student conflict with a student-driven intervention but that it matters which kind of students that get involved in the intervention; social referent students have bigger influence over social norms and behaviour in the school than other students have.

4.5 Spillover effects in neighborhoods and schools

In many other situations, there is also risk of interference between individuals. For example, interventions in schools might affect the pupils not directly receiving the treatment but being exposed to it through their classmates. Interventions in neighborhoods might affect residents not directly treated but influenced by their neighbors behaviour or decisions.

Hong and Raudenbush (2006) studied the effect of retaining low-achieving children in kindergarten instead of promoting them to first grade, on reading and math scale scores. If low-achieving children are retained it might affect the first grade class they otherwise would be in by facilitating teacher work when the first grade class is more homogeneous. Using the terminology by Ogburn and VanderWeele (2014), this is a typical example of *allocational interference* (see also Section 2).

As mentioned before, Lundin and Karlsson (2014) illustrated their proposed method, which is described in 3.1.2, using data on the effect of parent participation in a parenting program on their child's behaviour. The behaviour of untreated children might also change since they play together with treated children all day long in their preschool. This could be seen as an example of *interference by contagion* (cf. Section 2).

In one of the pioneering papers on causal inference under interference, Sobel (2006), the so called "Moving to Opportunity" (MTO) demonstration is given as an example of a situation when the no-interference assumption was not likely to hold but where many researchers have analyzed the data as if it did hold.

The MTO demonstration was a housing mobility experiment in five cities in the U.S., where volunteers living in high-poverty neighborhoods were randomly assigned to receive counseling and housing vouchers to move to low-poverty neighborhoods, or assigned to receive housing vouchers to move to any neighborhood, or assigned to a control group (for more details, see Sobel (2006) and references therein).

Due to the recruitment of MTO volunteers, by group meetings, it is very

likely that many participants knew other participants. If assigned to receive a treatment, an individual might more be likely to move if his/her friends and neighbors are also assigned to treatment and move than if he/she is the only one amongst the friends assigned to treatment. If this is the case, the no-interference assumption does not hold. Moreover, the no-interference assumption will not hold if, due to a tight rental market, the possibility to move is different if many individuals are assigned to treatment compared to if few individuals are assigned to treatment.

5 Discussion

In this paper we have reviewed a subset of statistical methods for causal inference under interference. These methods are important contributions to the field of causal inference, since the no-interference assumption (or SUTVA) is not plausible in many situations. It is only from 2012 and onwards the literature has begun to grow rapidly. There is also pioneering work from a few years earlier. Hudgens and Halloran (2008) is the most cited of these earlier papers and, in one way or the other, the starting point of a majority of the subsequent proposals.

We label the methods we included in the review as either methods for clustered data (Section 3.1) or methods for network data (Section 3.2). For clustered data there are suggestions for both randomized and non-randomized studies, while all proposals for network data, so far, are for randomized studies.

This paper does not give an absolute comprehensive summary of the field of causal inference under interference. As mentioned in Section 3, there are for example methods suggested for situations where data are paired. There are also papers arguing that there is a relation between interference and the seemingly unrelated area *causal interaction* which can be utilized so that the many existing empirical tests for causal interaction can be used to test for specific forms of interference (see VanderWeele et al. (2012)). Other papers suggesting tests for interference include Rosenbaum (2007), Aronow (2012), and Bowers et al. (2013). These papers were not included in our review since our focus was on estimation and not hypothesis testing. Moreover, the focus in this paper was on statistical methods suitable for labor market evaluations, which, in our eyes, disqualified, e.g., the methods for paired data.

It is difficult to evaluate the pros and cons of the suggested methods by comparing their performance with each other since they all are very situa-

tional, e.g., the estimands of interest are (very) different. The estimand in one study could be the ATEGT, which is the natural choice when considering if an intervention would be beneficial or not if applied to all individuals, in another study the natural choice of an estimand could be the k-level peer influence effect (Toulis and Kao, 2013) and in yet another study the population overall causal effect (Hudgens and Halloran, 2008) could be of main interest. Thus, there is no use in comparing properties such as the size of bias and variance of the estimators.

Worth noting is that all, but one, of "real world" examples covered in Section 4 assume partial interference, i.e., that all individuals in the study belongs to a group (cluster) and that there is no interference or spillover effects between individuals belonging to different groups. Moreover, of these examples, all but one (work absence) assume that the treatment assignment, regardless if it is randomized or not, is accomplished in two steps; first assigning "treatment strategies" to groups and then, conditional on the treatment strategy, assigning individuals within each group to treatment or control.

One drawback with these methods for causal inference under interference is the assumption of multiple and fixed groups; there are many situations where there are not enough groups (maybe there is just one group) or where the groups are not separated enough for partial interference to be a plausible assumption.

As far as we know, there are fewer examples of evaluations studies where any of the methods presented in Section 3.2 has been used. Among the examples in Section 4, only one (anti-conflict) concerns network data. This might of course be because these methods have not reached out to the practitioners yet, but it is probably because of lack of information about the network, i.e., which individuals in the study that socialize with each other. This information is typically not found in registers and is not, traditionally, asked for in questionnaires either. On the other hand, massive amount of information about peoples' online social networks are available to companies such as Facebook, Inc. and others. Many experiments on how, e.g., information and behavior spread within online networks have been done (see Centola, 2010; Bond et al., 2012). How such data could be used in the future is still unclear.

Conditionally on that it is possible to get information about how the network between individuals in the study is structured we think that all of the suggested methods presented in Section 3.2 have potential to be useful in evaluation studies, especially when indirect (spillover) effects are of main interest to measure. But the usefulness of the methods suggested by Toulis

and Kao (2013) and Aronow and Samii (2015) depend heavily on network topology of the network at hand. For example, due to network topology there might be very low probability for certain exposure classes making the variance of the IPW estimators very large (Aronow and Samii, 2015; Hellman and Lindberg, 2015). Thus, even if treatment assignment can be manipulated arbitrarily by the experimental design, the treatment exposure is constrained by the network topology.

Using the method suggested by Aronow and Samii (2015), i.e., estimating effects of treatment exposure instead of treatment effect, requires that the researcher has an idea about how the treatment is transmitted through the network. Is it only the direct neighbors that get exposed or is the treatment transmitted even further through the network? Is the exposure the same regardless of the number of treated neighbors and neighbors' neighbors and so on? This is a challenge of course, but also a possibility for the researcher; he/she can use several competing models for the treatment exposure mapping and maybe report estimates of treatment exposure effects from all of these.

All of the methods suggested for networks, so far, are for randomized studies. Since many evaluation studies of labor market programs are based on register data where program participation is not randomized to the unemployed, methods for non-randomized studies are much needed also for network data.

At a first glance, one could think that it should be possible to generalize the proposal of Aronow and Samii (2015) to non-randomized studies by using a set of observable covariates to estimate the probability to be assigned to treatment and then, from these estimated treatment probabilities and knowledge about the network, deduce the probability to be exposed to the different treatment exposure levels. For this to work, however, the set of covariates has to be such that conditional on them the potential outcomes should be independent of the treatment exposure level. This is different from the usual unconfoundedness assumption, i.e., that the potential outcomes are independent of the treatment assignment conditionally on the covariates. Thus, on a second thought, it might not be a simple task to generalize it to non-randomized studies, since this new type of unconfoundedness assumption might be difficult to assess.

We conclude this discussion with a quote from the last section in VanderWeele et al. (2014); "Much more exciting research is left to be done on causal inference under general forms of interference." In addition to this we want to point out that the remaining research is not only exciting but also very important. Ignoring interference can lead to seriously misleading

conclusions from evaluation studies. Methods that are viable in practise and with plausible assumptions are highly needed.

Acknowledgements

The authors wish to acknowledge The Institute for Evaluation of Labour Market and Education Policy for financial support. Also, the authors wish to thank the three reviewers for their valuable comments and suggestions.

References

Aronow, P. M. (2012). A general method for detecting interference between units in randomized experiments. *Sociological Methods & Research 41*(1), 3–16.

Aronow, P. M. and C. Samii (2015). Estimating average causal effects under interference between units. *arXiv:1305.6156v2 [math.ST]*, 1–35.

Bond, R. M., C. J. Fariss, J. J. Jones, A. D. Kramer, C. Marlow, J. E. Settle, and J. H. Fowler (2012). A 61-million-person experiment in social influence and political mobilization. *Nature 489*(7415), 295–298.

Bowers, J., M. M. Fredrickson, and C. Panagopoulos (2013). Reasoning about interference between units: A general framework. *Political Analysis 21*(1), 97–124.

Centola, D. (2010). The spread of behavior in an online social network experiment. *Science 329*(5996), 1194–1197.

Chiba, Y. (2012). A note on bounds for the causal infectiousness effect in vaccine trials. *Statistics & Probability Letters 82*(7), 1422–1429.

Crépon, B., E. Duflo, M. Gurgand, R. Rathelot, and P. Zamora (2013). Do labor market policies have displacement effects? evidence from a clustered randomized experiment. *The Quarterly Journal of Economics 128*(2), 531–580.

Eckles, D., B. Karrer, and J. Ugander (2014). Design and analysis of experiments in networks: Reducing bias from interference. *arXiv:1404.7530v2 [stat.ME]*, 1–32.

Ferracci, M., G. Jolivet, and G. J. van den Berg (2014). Evidence of treatment spillovers within markets. *Review of Economics and Statistics 96*(5), 812–823.

Gautier, P. A., P. Muller, B. van der Klaauw, M. Rosholm, and M. Svarer (2015). Estimating equilibrium effects of job search assistance. *CESifo Working Paper No. 5476*, 1–44.

Girma, S., Y. Gong, H. Görg, and S. Lancheros (2015). Estimating direct and indirect effects of foreign direct investment on firm productivity in the presence of interactions between firms. *Journal of International Economics 95*(1), 157–169.

Halloran, M. E. (2012). The minicommunity design to assess indirect effects of vaccination. *Epidemiologic methods 1*(1), 83–105.

Halloran, M. E. and M. G. Hudgens (2012). Causal inference for vaccine effects on infectiousness. *The international journal of biostatistics 8*(2), 1–40.

Halloran, M. E. and C. J. Struchiner (1991). Study designs for dependent happenings. *Epidemiology 2*(5), 331–338.

Halloran, M. E. and C. J. Struchiner (1995). Causal inference in infectious diseases. *Epidemiology 6*(2), 142–151.

Heckman, J. J., L. Lochner, and C. Taber (1999). Human capital formation and general equilibrium treatment effects: a study of tax and tuition policy. *Fiscal Studies 20*(1), 25–40.

Hellman, S. and E. Lindberg (2015). Interferens i kända och okända nätverk. Bachelor's thesis, Umeå University.

Hesselius, P., P. Johansson, and J. Vikström (2013). Social behaviour in work absence. *The Scandinavian Journal of Economics 115*(4), 995–1019.

Hesselius, P., J. P. Nilsson, and P. Johansson (2009). Sick of your collegues' absence? *Journal of the European Economic Association 7*(2-3), 583–594.

Hirano, K., G. Imbens, and G. Ridder (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica 71*(4), 1161–1189.

Hirano, K. and G. W. Imbens (2004). The propensity score with continuous treatments. In A. Gelman and X.-L. Meng (Eds.), *Applied Bayesian modeling and causal inference from incomplete-data perspectives*, pp. 73–84. Chichester: Wiley & Sons.

Hong, G. and S. W. Raudenbush (2006). Evaluating kindergarten retention policy. *Journal of the American Statistical Association 101*(475), 901–910.

Horvitz, D. G. and D. J. Thompson (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association 47*(260), 663–685.

Hudgens, M. G. and M. E. Halloran (2008). Toward causal inference with interference. *Journal of the American Statistical Association 103*(482), 832–842.

Johansson, P., A. Karimi, and J. P. Nilsson (2014). Gender differences in shirking: Monitoring or social preferences? evidence from a field experiment. *Working Paper 2014:9*.

Liu, L. and M. G. Hudgens (2014). Large sample randomization inference of causal effects in the presence of interference. *Journal of the American Statistical Association 109*(505), 288–301.

Lundin, M. and M. Karlsson (2014). Estimation of causal effects in observational studies with interference between units. *Statistical Methods & Applications 23*(3), 417–433.

Ogburn, E. L. and T. J. VanderWeele (2014). Causal diagrams for interference. *Statistical Science 29*(4), 559–578.

Paluck, E. L., H. Shepherd, and P. M. Aronow (2016). Changing climates of conflict: A social network experiment in 56 schools. *Proceedings of the National Academy of Sciences 113*(3), 566–571.

Rigdon, J. (2015). *interferenceCI: Exact Confidence Intervals in the Presence of Interference*. R package version 1.1.

Rigdon, J. and M. G. Hudgens (2015). Exact confidence intervals in the presence of interference. *Statistics & probability letters 105*, 130–135.

Rosenbaum, P. R. (2007). Interference between units in randomized experiments. *Journal of the American Statistical Association 102*(477), 191–200.

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

Rubin, D. B. (1980). Discussion of paper by D. Basu. *Journal of the American Statistical Association 75*, 591–593.

Rubin, D. B. (2010). Reflections stimulated by the comments of Shadish (2010) and West and Thoemmes (2010). *Psychological Methods 15*(1), 38–46.

Sobel, M. (2006). What do randomized studies of housing mobility demonstrate? *Journal of the American Statistical Association 101*(476), 1398–1407.

Tchetgen Tchetgen, E. J. and T. J. VanderWeele (2012). On causal inference in the presence of interference. *Statistical Methods in Medical Research 21*(1), 55–75.

Toulis, P. and E. Kao (2013). Estimation of causal peer influence effects. In *Proceedings of The 30th International Conference on Machine Learning*, pp. 1489–1497.

van der Klaauw, B. (2014). From micro data to causality: Forty years of empirical labor economics. *Labour Economics 30*, 88–97.

VanderWeele, T. J., E. J. Tchetgen Tchetgen, and M. E. Halloran (2014). Interference and sensitivity analysis. *Statistical Science 29*(4), 687–706.

VanderWeele, T. J., J. P. Vandenbroucke, E. J. Tchetgen Tchetgen, and J. M. Robins (2012). A mapping between interactions and interference: implications for vaccine trials. *Epidemiology 23*(2), 285–292.

Verbitsky-Savitz, N. and S. W. Raudenbush (2012). Causal inference under interference in spatial settings: A case study evaluating community policing program in chicago. *Epidemiologic Methods 1*(1), 107–130.