THREE ESSAYS ON THE INFLUENCE OF PEERS AND PRIMARY CARE ENGAGEMENT

A Dissertation
Submitted to
the Temple University Graduate Board

In Partial Fulfillment of the
Requirements for the Degree
DOCTOR OF PHILOSOPHY

by
Edward Patrick Kost
August 2023

Examining committee members:

Johanna Catherine Maclean, Advisory Chair, Department of Economics
Dimitrios Diamantaras, Department of Economics
Douglas Webber, Department of Economics
Edward Rosenthal, Department of Statistics and Data Science

ABSTRACT

In this dissertation, I study econometric issues in network and health economics.

Measurement error is a ubiquitous problem in the peer effects literature that is not well understood. In Chapter 1, "Measurement error in peer effects," I develop a constructive approach to empirically assess the bias caused by links missing at random. I apply my method to study the bias in peer effect estimates of recreational and physical activities among adolescents in the United States. I find that the magnitude and direction of the bias depends on the estimator. Estimators that measure the aggregate effect of peers' outcomes are more robust to measurement error and can be unbiased even when fifty percent of peer interactions are unobserved. Estimators that measure the average effect of peers' outcomes are more susceptible to measurement error and suffer from a persistent downward bias. My findings illustrate the importance of understanding measurement error's impact, when measurement error will likely bias results and when it can be safely ignored.

In Chapter 2, "Non-random errors in peer effects," I study the effects of measurement error on a generalized peer effect model that nests two of the most commonly used estimators. Measurement error in the specification of peer groups leads to biased estimates. I adapt Monte Carlo methods developed for studying measurement error when peers' interactions are missing at random to understand the effects of top-coding, non-random errors and spurious peer interactions. I find that non-random errors pose the greatest threat, often leading to overestimation

and persistent biases. Top-coding can also severely bias estimates when the constraint impacts a majority of individuals but otherwise has a mild effect. While spurious links in limited quantities can often be ignored.

Chapter 3, "Nurse outreach and frequent emergency department users: A synthetic control analysis," studies the effects of an intervention to promote primary care engagement among frequent emergency department users. Emergency departments are one of the costliest places to receive care and are routinely overcrowded. Various policy initiatives have yielded mixed findings. I use synthetic control methods to analyze the effects of a nurse outreach program for frequent emergency department users implemented by a major U.S. insurer. The program seeks to reduce emergency department utilization by promoting primary care engagement. I leverage a unique commercial claims data set to measure the effects of the program on primary care and emergency department utilization. My findings suggest that six months after treatment nurse outreach increased primary care utilization by 15 percent; however, I find no clear effect on emergency department utilization. My findings indicate that increasing primary care engagement may not be sufficient to prevent emergency department over utilization.

DEDICATION

To my family, thank you for your love and support throughout this journey.

ACKNOWLEDGEMENTS

I would like to thank my committee Catherine Maclean, Dimitrios Diamantaras, Douglas Webber and Edward Rosenthal for their guidance, comments and support. I would also like to thank Charles Swanson and the Graduate Economics Seminar participants for their feedback and suggestions. This research uses data from Add Health, a program project designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris, and funded by a grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 17 other agencies. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Persons interested in obtaining Data Files from Add Health should contact Add Health, The University of North Carolina at Chapel Hill, Carolina Population Center, Carolina Square, Suite 210,123 W. Franklin Street, Chapel Hill, NC 27516 (addhealth contracts@unc.edu). No direct support was received from grant P01-HD31921 for this analysis. All data was provided on an as-is basis. All errors are my own.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

MEASUREMENT ERROR IN PEER EFFECTS

1.1 Introduction

A growing literature documents a strong correlation between the outcomes of individuals and that of their peers (Bramoullé et al., 2020). Risk taking behavior, educational outcomes, financial decisions, even the likelihood of being published in a top economics journal all seem to depend on these underlying social structures (Halliday and Kwak, 2012; Ammermueller and Pischke, 2009; Sojourner, 2013; Banerjee et al., 2021; Carrell et al., 2022). Yet, estimating the importance of peer effects requires econometricians to overcome a number of challenges.

A prime example arises when trying to distinguish between the effects of peers' characteristics and that of their outcomes. The outcomes of interacting peers' are often observed simultaneously. As a result, peer effect estimates can suffer from a form of simultaneity bias known as the "reflection problem" (Manski, 1993). The reflection problem can be resolved when individual specific peer groups are observable as shown by Bramoullé et al. (2009), De Giorgi et al. (2010), Liu and Lee (2010) and Laschever (2011).

However, as De Giorgi et al. (2010) point out, observing such interactions presents its own challenges. The observed network may differ from the true network because of the sampling method, top-coding, network boundary constraints, mis-coding, mis-reporting or non-response (Advani and Malde, 2018). Beyond that, there is an even more fundamental problem. The observed network may not be the relevant one. The relevant network of interactions depends on the social mechanism that links peers' outcomes. Identifying that mechanism is not an econometric problem, and more often than not the potential impact of measurement error is swept under the rug. For all of these reasons, prior researchers have cautioned that measurement error is a serious and ubiquitous problem that warrants further research (De Giorgi et al., 2010; Chandrasekhar and Lewis, 2011; Bramoullé et al., 2020).

While measurement error can in some cases be addressed analytically, often there is no known remedy. Instead several researchers have used Monte Carlo simulations to estimate the impact of measurement error (De Giorgi et al., 2010; Liu, 2013; Griffith, 2019). This approach involves comparing the observed network with a large number of counterfactuals.

In this paper, I propose a new method for generating counterfactual networks that can cover a broader range of scenarios with fewer computations when compared with traditional methods. The core of my approach relies on the fact that own-outcomes depend additively on peers' outcomes. As a result, very little changes when only one peer interaction is added to (or removed from) the network. By starting with a network with no interactions and varying the order in which interactions are added, one can map out the bias caused by an entire set of misspecifications.

I present my method in the context of a generalized peer effects model due to Liu et al. (2014). The model nests the local-average model of Bramoullé et al. (2009) and the local-aggregate model of Liu and Lee (2010). I illustrate through a simple example how misspecified peer groups violate the relevance and exclusion restrictions placed on the model and lead to biased results.

I then demonstrate how my method can be applied to study the effects of misspecified peer groups when links are missing at random. I illustrate with a simple running example how each of the data elements in my model evolve as links are added. I show that the expected number of computations needed for my method is strictly smaller than that of the stochastic methods used by prior researchers.

I then apply my methodology to data from the National Longitudinal Study of Adolescent to Adult Health (Add Health) and a series of simulated networks. When links are missing at random, the local-aggregate effect of peers' outcomes exhibits only a modest inflationary bias even when the majority of links are unobserved with little variation in estimates. Conversely, the local-average effect has a strong negative bias when peer groups contain many missing relationships, and the effect is estimated imprecisely even when almost the entire network is observed. Although more research is needed to understand the effects of other forms of measurement error, knowing the effects of links missing at random contributes to our understanding of the precision of peer effect models.

The rest of the paper is organized as follows. Section 2 defines the notation and empirical framework used throughout the paper. Section 3 reviews the literature on measurement error in peer effects. Section 4 describes the algorithm. Section 5 discusses the empirical application. Section 6 analyzes the results, and section 7 concludes.

1.2 Empirical Framework

In this section I introduce the notation and empirical model used throughout the paper. I consider a network of n interacting individuals. Denote by G a $n \times n$ adjacency matrix that records those interactions with typical element $g_{ij} = 1$ if individual i is linked to individual j, and 0 otherwise.[1] Denote the neighborhood of i, that is the set of individuals that i is linked to by $n(i)$, with size $g_i = |n(i)|$. Finally, define the row-normalized adjacency matrix A as the $n \times n$ matrix obtained by dividing each of the rows of G by g_i, so that $a_{ij} = g_{ij}/g_i$.

Given these definitions, I consider the following generalized peer effects model based on Liu et al. (2014)

$$y_{i,r} = \lambda_1 \sum_{j \in n(i,r)} y_{j,r} + \lambda_2 \frac{1}{g_{i,r}} \sum_{j \in n(i,r)} y_{j,r} + \mathbf{x}_{i,r}\beta + \frac{1}{g_{i,r}} \sum_{j \in n(i,r)} \mathbf{x}_{j,r}\gamma + \delta_r + \epsilon_{i,r}, \quad (1.1)$$

that is $y_{i,r}$, the observed outcome of individual i in cluster r, depends on the sum of their neighbors' outcomes, $\sum_{j \in n(i,r)} y_{j,r}$, the average of their neighbors' outcomes, $\frac{1}{g_{i,r}} \sum_{j \in n(i,r)} y_{j,r}$, their own-characteristics, $\mathbf{x}_{i,r}$, the average of their neighbors' characteristics, $\frac{1}{g_{i,r}} \sum_{j \in n(i,r)} \mathbf{x}_{j,r}$, a cluster fixed-effect δ_r and an idiosyncratic shock $\epsilon_{i,r}$.

For example, in Bramoullé et al. (2009) where y_{ir} measures the number of recreational activities that an individual participates in, when $\lambda_1 > 0$, own recreational activity level will tend to be high whenever the aggregate activity level of the neighborhood is high, regardless of how those activities are distributed among neighbors, and when $\lambda_2 > 0$, own activity will tend to be high only when the average level of neighbors' activity is high.

[1] As is standard in the literature, I assume that individuals may not interact with themselves so that $g_{ii} = 0$ for all i.

The parameters, β and γ, control for the influence of own-characteristics and of neighbors' exogenous characteristics on own-outcomes. While, the fixed effect δ_r captures unobserved cluster-level shocks. The term cluster refers to a feature of the network topology. Specifically, a cluster is a maximal subset of individuals such that every individual in a given cluster can be reached by every other member of the same cluster by traversing some sequence of links in either direction. Clusters are a method of identifying groups of individuals that share common experiences based on the network structure. When the number of clusters is large, it is common to apply a within transformation to eliminate the cluster fixed-effects and avoid the incidental parameters problem.

Researchers are primarily interested in estimates of λ_1 and λ_2, often called the local-aggregate and local-average effects because they depend on aggregate and average neighborhood outcomes, respectively. When λ_1 is positive equation (2.1) is said to exhibit a social multiplier, where individual policy shocks can have repercussions for the behavior of the entire network. Conversely, when λ_2 is positive there is a greater tendency towards social conformity. Deviations from the social norm can be costly, and individual shocks tend to dissipate. Policies that seek to affect the entire network must change the social norm to be effective.

1.2.1 A note on notation

In equation (2.1) I express $y_{i,r}$ as a function of neighborhoods in order to highlight the central role that neighbors play in my analysis. However, equation (2.1) can just as easily be written in matrix notation. I make use of this fact throughout the rest of the paper in order to simplify certain concepts that do not translate into the neighborhood notation easily. As such, readers should be aware of the following identities. The vector of the sums of neighbors' outcomes can be written simply as GY, where

$$GY = \left\{ \sum_{j \in n(1)} y_j, \sum_{j \in n(2)} y_j, \ldots, \sum_{j \in n(n)} y_j \right\}', \tag{1.2}$$

and equality follows from the definitions of matrix multiplication, the adjacency matrix, G, and neighborhoods, $n(\cdot)$. Similarly, the vector of the averages of neighbors' outcomes can be represented as,

$$AY = \left\{ \frac{1}{g_1} \sum_{j \in n(1)} y_j, \frac{1}{g_2} \sum_{j \in n(2)} y_j, \ldots, \frac{1}{g_n} \sum_{j \in n(n)} y_j \right\}', \tag{1.3}$$

where equality follows from the definition of matrix multiplication and of the row-normalized adjacency matrix, A, that is $a_{ij} = g_{ij}/g_i$. More broadly, G (or A) times any conformable matrix Z can be expressed as a function of neighborhoods, where GZ gives the sum of neighbors' z_j and AZ the average.

1.2.2 Estimation

The parameters in equation (2.1) cannot usually be estimated by ordinary least squares. Manski (1993) shows that when using an entire classroom as the peer network the average of neighbors' outcomes and the average of neighbors' characteristics become perfectly collinear in expectations. This is the "reflection problem," a form of simultaneity bias.

Liu et al. (2014) prove that the local-aggregate and local-average effects can be identified from equation (2.1) given the linear independence of I, G, A, GA, A^2, GA^2 and A^3 and that either $\lambda_1\beta \neq 0$ or $\lambda_2\beta + \gamma \neq 0$. The first condition ensures that there is sufficient variation in the neighborhoods of connected agents to identify the model parameters. Without the linear independence assumption the instruments and regressors may be perfectly collinear and identification fails.

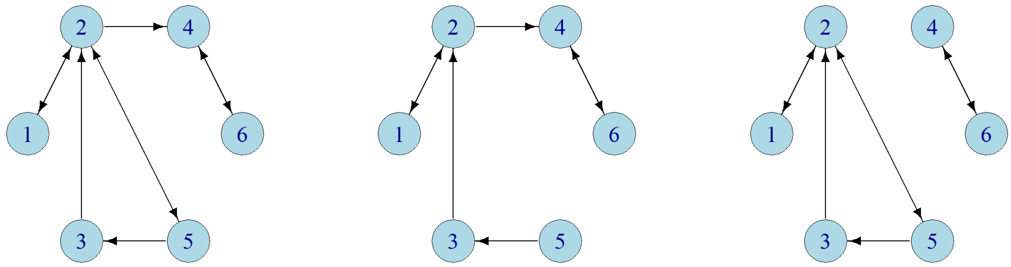
The second condition assumes that either own characteristics or neighbors characteristics are relevant predictors. If there are no relevant exogenous predictors, then there are no valid instruments for identification. These conditions will typically hold whenever neighborhoods vary at the level of the individual and relevant exogenous regressors are observed.

When the above assumptions are met the sum of neighbors' outcomes can be instrumented for with neighborhood size, g_i, and the sum of neighbors' characteristics, $\sum_{j\in n(i)} \mathbf{x}_j$. Similarly, the average of neighbors' outcomes can be instrumented for with the average of neighbors' neighbors' average characteristics, i.e. A^2X. The intuition behind using A^2X as an instrument follows from the fact that the non-zero elements of A^2 depend on neighbors' neighbors'. Suppose that there exists a triad, $\{i, j, k\}$, such that individuals i and k are not neighbors with one another but share a mutual neighbor j. Then the outcomes of i and k will be correlated because they both depend on j, but they will not influence each other directly because they are not neighbors. Such triads are referred to as *open* triads. If $\{i, j, k\}$ were all neighbors with one another, they would be referred to as a closed triad. Open triads satisfy the relevance and exclusion restrictions for a set of potential instruments. However, the validity of these instruments implicitly assumes that the network is observed without error.

1.3 Measurement error

Blume et al. (2015) show that when the adjacency matrix, G, is measured with error models like equation (2.1) are biased. To illustrate this point consider the example network depicted in Figure 1.1 panel (a). The network contains a single cluster of six individuals and nine links. Six open triads, $\{1, 2, 4\}$, $\{1, 2, 5\}$, $\{3, 2, 4\}$, $\{5, 2, 1\}$, $\{5, 2, 4\}$ and $\{2, 4, 6\}$, form sets of potential instruments for the average of neighbors' outcomes. Now, consider what happens when this network

Figure 1.1
Examples of measurement error



(a) (b) (c)

Panel (a) shows a simple network used as an example throughout the text to illustrate various ideas. Panels (b) and (c) illustrate possible mis-specifications of the network in panel (a). In panel (b) the (2,5) and (5,2) links are unobserved violating the exclusion restriction for potential instruments. In panel (c) the (2,4) link is unobserved leading to a less informative set of instruments.

is observed with error. Suppose, as depicted in Figure 1.1 panel (b), the links between 2 and 5 are not observed, e.g. because these two individuals forgot to report that they are friends, creating a new open triad $\{5, 3, 2\}$. Then 2 would appear to be a valid instrument for 5 when in fact the exclusion restriction is violated because 2 and 5 influence each other directly. Alternatively, suppose that 2 failed to report that they are friends with 4, as depicted in Figure 1.1 panel (c). This eliminates four of the open triads. The resulting instrument is far less informative.

Although no analytical solution exists to address missing links as described above, De Giorgi et al. (2010) provide some early evidence on their effect using Monte Carlo simulations. In the authors primary analysis, college students' choice of major is modeled as a function of their classmates' choices. Distinct peer groups are identified by leveraging differences in class assignment.

The authors recognize that the class network may not be the relevant one. They explore the bias associated with estimates based on class assignment when the relevant peer group is friends. In order to make their model tractable the authors assume that class assignment is random and friendship depends probabilistically on class assignment. The authors simulate their model by stochastically changing the observed class assignment network to obtain counterfactual friendship networks.

After simulating many class assignment and friendship network pairs and comparing estimates based on the misspecified class networks, De Giorgi et al. (2010) conclude that the bias in their simulations can be either inflationary or attenuating but is generally small. Although their results are encouraging, their results suffer from several limitations. In their simulations, errors can arise either because individuals become friends outside of class or because individuals fail to become friends in class. The authors only consider a handful of scenarios, essentially a high and low probability of each type of error occurring. These are two different error processes with potentially opposing effects. When individuals become friends outside of class the class assignment network fails to observe the link reducing neighborhood size, g_i, but when individuals fail to become friends in class, the class assignment network "observes" a link that should not be there and neighborhood size is inflated. Modeling both types of errors simultaneously conflates their effects.

Liu (2013) and Griffith (2019) use similar Monte Carlo methods to estimate the magnitude of sampling biases. Sampling bias is perhaps the best understood form of network measurement error. Both because sampling induced errors occur in every study that fails to complete a full census and because it can be relatively straightforward to correct for sampling induced errors analytically, under the right circumstances.

Chandrasekhar and Lewis (2011) and Liu (2013) study the implications of using two common sampling methods. The first method supposes that researchers have access to information on all n individuals in a population, e.g. a school roster with administrative data. The researchers survey s individuals and ask each respondent to nominate their friends from the remaining $n-1$ individuals. In the second method, researchers only possess information on the s sampled individuals and ask respondents to nominate friends from among the other $s-1$ individuals. Chandrasekhar and Lewis (2011) show that the divergence between the true and observed networks caused by either method results in an errors-in-variables problem and non-classical errors.

However, they also propose a simple analytical correction for the first sampling method. The authors observe that in the first method the sum of neighbors' outcomes and the averages of neighbors' outcomes and characteristics, that is $\sum_{j\in n(i)} y_j$, $\frac{1}{g_i}\sum_{j\in n(i)} y_j$ and $\frac{1}{g_i}\sum_{j\in n(i)} \mathbf{x}_j$, are unbiased for sampled individuals because y_i and x_i are observed for everyone in the population. Thus, there is no measurement error in a second stage estimate of sampled individuals' neighborhood characteristics, and the error in the first stage is uncorrelated with the second stage residuals satisfying the exclusion restriction.

Liu (2013) employ Monte Carlo methods to show that estimates based on the first sampling method can still be biased when the sampling rate becomes sufficiently low because the first stage regression ceases to be informative, but Liu (2013) also demonstrate that unbiased estimates of the local-aggregate effect, λ_1, can still be obtained even for low sampling rates. For sampled individuals the first sampling method still provides unbiased estimates of neighborhood size, g_i, regardless of the sampling rate. Thus, the first stage regression remains informative for λ_1 even when sampling rates are low.

Sampling can also generate biased results if respondents are limited in the number of individuals they can be linked to, what is known as top-coding. Griffith (2019) analyzes the effects of top-coding in a model with only exogenous peer effects. Using Monte Carlo simulations he finds that top-coding can significantly bias peer effect estimates. Under the assumption that the order in which respondents report peers is irrelevant Griffith (2019) shows that the expected bias is attenuating. He also provides bias correction methods for estimating the exogenous effect, γ, that like Liu (2013) rely on a consistent estimate of neighborhood size.

1.4 Methods

My method builds on the Monte Carlo simulations of De Giorgi et al. (2010), Liu (2013) and Griffith (2019), but rather than stochastically altering the network in equation (2.1), I consider how estimates of equation (2.1) evolve as the network changes one link at a time.

To fix ideas, imagine trying to estimate the impact of links missing at random on a given network. One could select a fixed-size random sample of non-zero elements from the adjacency matrix, G, set those elements equal to zero and then derive the remaining data elements needed to estimate equation (2.1) from the misspecified matrix. The differences between parameter estimates based on the original network and the misspecified network measure the bias induced by the missing links. Repeating this process, each time choosing an equally sized, independently random sample of non-zero elements from G, would allow one to characterize the empirical distribution of the bias induced by that fraction of missing links. This is the basic idea behind the methods employed by De Giorgi et al. (2010), Liu (2013) and Griffith (2019).[2]

Deriving the data elements necessary to estimate equation (2.1) from the misspecified adjacency matrix requires multiplying the misspecified G times, Y, X

[2]De Giorgi et al. (2010) not only set a fraction of the ones in G equal to zero, they also set a fraction of the zeroes equal to one. Liu (2013) simulate sampling by removing entire rows from G. Griffith (2019) simulate top-coding by selecting non-zero elements of the adjacency matrix in a way that bounds the row sums of G, thereby limiting neighborhood size.

Figure 1.2
Example evolution of links

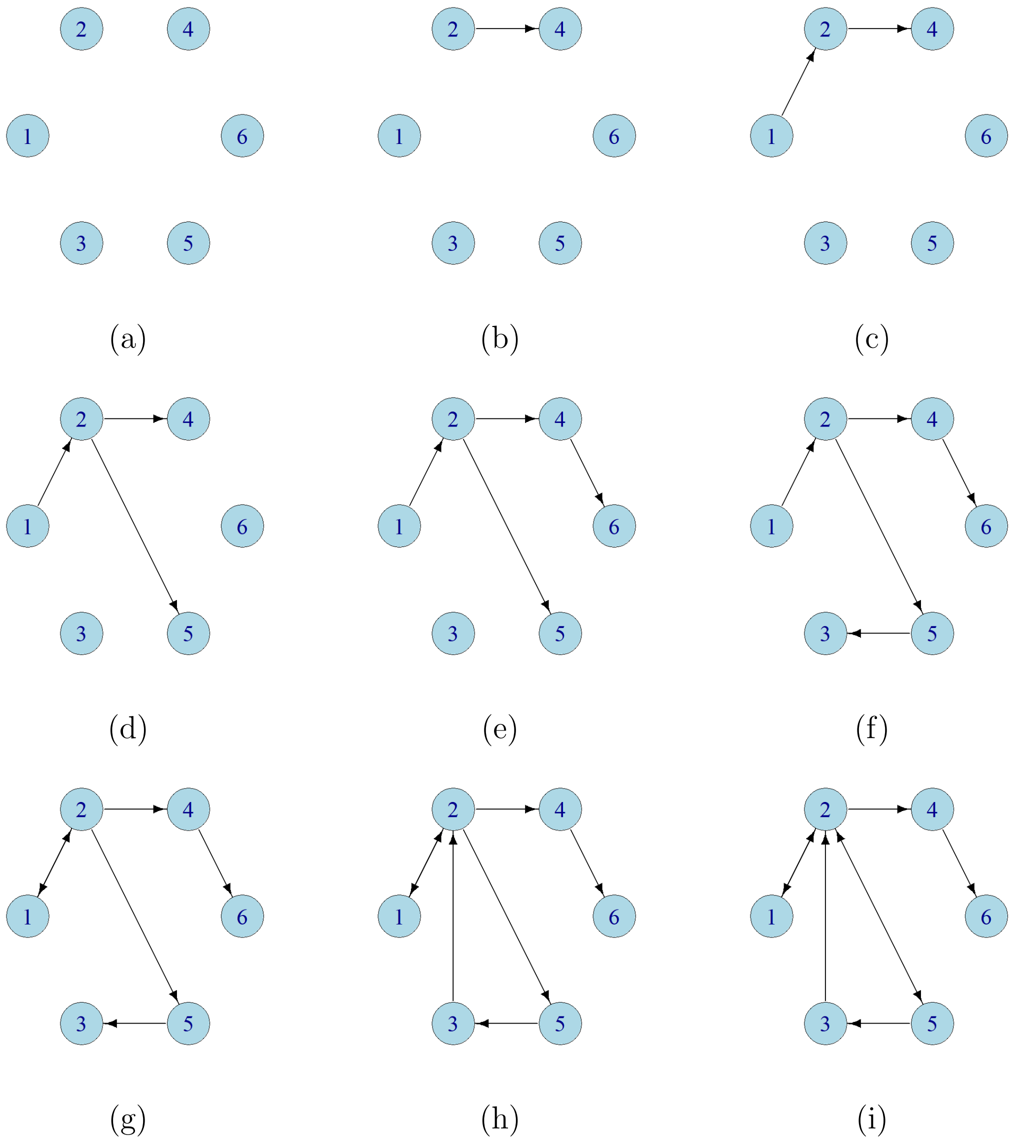


Figure 1.2 continues with the simple network introduced in Figure 1.1 panel (a). Figure 1.1 panel (a) shows the empty network, that is the simple network with all links removed. Panels (b) - (i) illustrate adding the links from the original graph in the order described by the ordered tuple $\ell = \{(2,4), (1,2), (2,5), (4,6), (5,3), (2,1), (3,2), (5,2), (6,4)$.

and so on. Since G is typically sparse an efficient matrix multiplication algorithm can calculate GY in a number of steps that is proportional to the number of non-zero elements in G, i.e. the number links.

The core insight of my method is that each step in the sparse matrix multiplication algorithm is analogous to adding one link to the network. By adding links one at a time, one can enumerate the changes caused by each additional link. Furthermore, by adding links in the "right" order, one can model links that are missing at random. The number of calculations in my method remains proportional to the number of links, but because my method enumerates the changes caused by each added link, I eliminate the need for repeated simulations when studying different fractions of missing links.

1.4.1 Details

My approach entails the following steps. Start with a given network. Remove all of the links from that network and store them in an ordered tuple. Call the network with no links, the empty network, and in a slight abuse of notation the ordered tuple as well as its length simply as ℓ.

Arrange ℓ so that links are ordered uniformly at random. Begin adding the links in ℓ back to the empty network one at a time. See Figure 1.2 panels (a) through (l) for an illustration based on the example network from Figure 1.1 with $\ell = \{(2,4), (1,2), (2,5), (4,6), (5,3), (2,1), (3,2), (5,2), (6,4)\}$.

After the addition of k links, the resultant network can be thought of as an observation of the original network with $\ell - k$ missing links. Furthermore, because the links in ℓ are ordered uniformly at random, the missing links are missing at random. The difference between an estimate of equation (2.1) based on the original network and the network with k links is an estimate of the bias induced by observing the original network with $\ell - k$ links missing at random. Comparing estimates for each value of k from 1 to ℓ enumerates ℓ realizations of that bias. Repeating this process many times, each with a new uniformly random order for ℓ, produces an empirical distribution that characterizes the bias caused by the missing links.[3] The rest of this section describes how the data elements in my model change as links are added as well as the computational complexity of those changes.

1.4.2 Tracking neighborhoods

Recall that the instrumental variables estimator described in Section 2 involves observations of Y, X, GY, AY, AX and a cluster indicator as well as the instruments g, GX and A^2X. Assume that Y and X are known, e.g. because of the availability of a school roster and administrative data, then given a sequence of links, ℓ, I can describe the evolution of the remaining data elements.

To illustrate how the data elements evolve, I use the example network from Figure 1.1 and the uniformly random sequence of links, $\ell = \{(2,4),\ (1,2),\ (2,5),\ (4,6),\ (5,3),\ (2,1),\ (3,2),\ (5,2),\ (6,4)\}$, from Figure 1.2 as a running example throughout this section. The empty network in Figure 1.2 panel (a) has n empty neighborhoods, so that

$$n(1) = n(2) = \cdots = n(6) = \{\emptyset\}.$$

Figure 1.2 panel (b) depicts adding the first link, $(2,4)$, from ℓ to the empty network. The addition of link $(2,4)$ to the empty network joins individual 4 to the neighborhood of individual 2, so that $n(2) = \{\emptyset\} \cup \{4\} = \{4\}$, while the remaining $n-1$ neighborhoods are unaffected. Adding the next link, $(1,2)$, depicted in Figure

[3] In practice, it is usually sufficient to estimate the bias for values of k at pre-determined intervals, e.g. $k \in \{0.10|\ell|, 0.20|\ell|, \ldots, 0.90|\ell|\}$, rather than every value of k. This avoids the computational burden of estimating equation (2.1) $|\ell|$ times while providing sufficient information to describe the bias.

1.2 panel (c), joins 2 to the neighborhood of 1, so that $n(1) = \{\emptyset\} \cup \{2\} = \{2\}$. Adding, $(2, 5)$, as in Figure 1.2 panel (d), gives $n(2) = \{4\} \cup \{5\} = \{4, 5\}$, and so on. Each new link always leads to one element being added to one neighborhood. Therefore, updating the set of neighborhoods requires one operation per link.

Updating the vector of neighborhood sizes also requires one operation per link. The size of neighborhood i, that is g_i, is just the number of elements in that neighborhood. Thus, each new link, (i, j), simply increments g_i by one. Continuing with the example network above, the empty network is such that

$$g = \{0, \dots, 0\}' . \tag{1.4}$$

The addition of link $(2, 4)$ to the empty network increases g_2 by one, so that $g = \{0, 1, 0, \dots, 0\}$. After adding link $(1, 2)$, g_1 increases by one, so that $g = \{1, 1, 0, \dots, 0\}$, and after adding $(2, 5)$, g_2 increases again by one, so that $g = \{1, 2, 0, \dots, 0\}$, and so on.

The evolution of the sums of neighbors' outcomes, GY, also needs only one operation per link, but now each link, (i, j), increments the i^{th} sum in GY by y_j. When the network is empty

$$GY = \left\{ \sum_{j \in n(1)} y_j, \sum_{j \in n(2)} y_j, \dots, \sum_{j \in n(n)} y_j \right\}' = \{0, \dots, 0\}' \tag{1.5}$$

because each of the n neighborhoods are empty and the sum of the numbers in the empty set equals zero by convention. In the toy network, after the addition of link $(2, 4)$, $GY = \{0, y_4, 0, \dots, 0\}'$. After adding link $(1, 2)$, $GY = \{y_2, y_4, 0, \dots, 0\}'$, and after adding $(2, 5)$, $GY = \{y_2, y_4 + y_5, 0, \dots, 0\}'$, and so on.

The evolution of the sums of neighbors' characteristics, GX, extends this idea to a matrix with p columns and as a result requires p calculations per link. When

X is $n \times p$, adding link $(2,4)$ to the toy network above yields

$$GX = \left\{ \begin{array}{cccc} 0 & 0 & \cdots & 0 \\ x_{4,1} & x_{4,2} & \cdots & x_{4,p} \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 \end{array} \right\}, \tag{1.6}$$

where the first subscript indicates the row of X and the second subscript indicates the column. After adding link $(1,2)$,

$$GX = \left\{ \begin{array}{cccc} x_{2,1} & x_{2,2} & \cdots & x_{2,p} \\ x_{4,1} & x_{4,2} & \cdots & x_{4,p} \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 \end{array} \right\}, \tag{1.7}$$

and after adding $(2,5)$,

$$GX = \left\{ \begin{array}{cccc} x_{2,1} & x_{2,2} & \cdots & x_{2,p} \\ x_{4,1} + x_{5,1} & x_{4,2} + x_{5,2} & \cdots & x_{4,p} + x_{5,p} \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 \end{array} \right\}. \tag{1.8}$$

The evolution of the averages of neighbors' outcomes and characteristics, AY and AX, are described entirely by the evolutions of the sizes of neighborhoods, g, and the sums of neighbors' outcomes and characteristics, GY and GX. To see this, compare the representations of GY and AY from equations (1.2) and (1.3),

$$GY = \left\{ \sum_{j \in n(1)} y_j, \sum_{j \in n(2)} y_j, \ldots, \sum_{j \in n(n)} y_j \right\}'$$

and

$$AY = \left\{ \frac{1}{g_1} \sum_{j \in n(1)} y_j, \frac{1}{g_2} \sum_{j \in n(2)} y_j, \ldots, \frac{1}{g_n} \sum_{j \in n(n)} y_j \right\}'.$$

From these representations it should be clear that the i^{th} row of AY equals the i^{th} row of GY divided by g_i, and by extension the i^{th} row of AX equals the i^{th} row of GX divided by g_i. Thus, knowledge of g, GY and GX is sufficient to describe AY and AX.

The evolution of the instrument A^2X is a bit more complicated. The following observations allow the problem to be broken down into more manageable parts. First, recall that A times any conformable matrix, Z, can be represented as n neighborhood averages. Second, note that because matrix multiplication is associative, I can write $A^2X = A(AX) = AZ$, where $Z = AX$. Thus, the evolution of A^2X can be thought of as resulting from two components, the evolution of AX and of $A(AX)$.

To illustrate this point let us return to the example network in 1.2. The addition of link $(2, 4)$ to the empty network adds the fourth row of X to the average in the second row of AX, so that

$$AX = \left\{ \begin{array}{cccc} 0 & 0 & \cdots & 0 \\ x_{4,1} & x_{4,2} & \cdots & x_{4,p} \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 \end{array} \right\}, \tag{1.9}$$

and adds the fourth row of AX to the average in the second row of A^2X. Since the fourth row of AX equals zero, adding link $(2,4)$ trivially yields,

$$A^2X = \begin{Bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{Bmatrix}. \tag{1.10}$$

The addition of link $(1,2)$ adds the second row of X to the average in the first of row of AX, so that

$$AX = \begin{Bmatrix} x_{2,1} & x_{2,2} & \cdots & x_{2,p} \\ x_{4,1} & x_{4,2} & \cdots & x_{4,p} \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 \end{Bmatrix}, \tag{1.11}$$

and adds the second row of AX to the average in the first of row of A^2X. Thus, adding link $(1,2)$ yields

$$A^2X = \begin{Bmatrix} x_{4,1} & x_{4,2} & \cdots & x_{4,p} \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{Bmatrix}. \tag{1.12}$$

Adding link $(2,5)$ adds the fifth row of X to the average in the second row of AX, so that

$$AX = \begin{Bmatrix} x_{2,1} & x_{2,2} & \cdots & x_{2,p} \\ \frac{1}{2}(x_{4,1}+x_{5,1}) & \frac{1}{2}(x_{4,2}+x_{5,2}) & \cdots & \frac{1}{2}(x_{4,p}+x_{5,p}) \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 \end{Bmatrix}, \tag{1.13}$$

and adds the fifth row of AX to the average in the second row of A^2X. Since the fifth row of AX equals zero the second row of A^2X remains unchanged. However, the first row of A^2X depends on the second row of AX and the second row of AX has changed. Therefore, the first row of A^2X must also change to reflect the new value of AX. After adding link $(2,5)$

$$A^2X = \begin{Bmatrix} \frac{1}{2}(x_{4,1}+x_{5,1}) & \frac{1}{2}(x_{4,2}+x_{5,2}) & \cdots & \frac{1}{2}(x_{4,p}+x_{5,p}) \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{Bmatrix}. \tag{1.14}$$

The above example illustrates how adding a link, (i,j), can change not only the i^{th} row of A^2X but any row, k, of A^2X, such that i is in the neighborhood of k. These two effects are a result of the evolution of neighborhoods and of the average of neighbors characteristics. Fortunately, only a little extra bookkeeping is necessary to track these effects.

To that end, define $\hat{n}(i)$ as the set of individuals whose neighborhoods include i. Then the addition of link, (i,j), entails adding i to the set of individuals whose neighborhoods include j, that is $\hat{n}(j)$, in addition to adding j to the neighborhood of i. In essence, it is necessary to track both ends of the link.

The effected rows of A^2X are then given by the union of i and $\hat{n}(i)$. Row i because of the evolution of neighborhood i and the rows of AX that depend on i. The number of calculations needed to track the changes in A^2X is therefore proportional to $1 + |\hat{n}(i)|$. Since the number of elements in the set $\{\hat{n}(1), \hat{n}(2), \ldots, \hat{n}(n)\}$ grows at the same rate as the vector of neighborhood sizes, the number of calculations will be proportional to average neighborhood size.

The last object that needs to be addressed in order to estimate equation (2.1) is the matrix of cluster indicators. Recall that a cluster is defined as a maximal subset of individuals such that every individual in a given cluster can be reached by every other member of the same cluster by following some sequence of links in either direction. Newman and Ziff (2001) provide an efficient method to track cluster membership and size when adding links one at a time starting with an empty network. Here I outline an easy to implement version of their algorithm that captures the central ideas necessary to proceed.[4]

The cluster tracking algorithm works as follows. Define two vectors of length n. The first vector, m, will be used to track cluster membership. The second vector, s, will be used to track cluster size. For an empty network of size n, $m = \{1, 2, \ldots, n\}$ and $s = \{1, 1, \ldots, 1\}$, so that each individual belongs to their own cluster of size one. After the addition of link (i, j) first compare m_i to m_j. If $m_i = m_j$, then stop, i and j already belong to the same cluster. If $m_i \neq m_j$, then compare the sizes of each cluster, i.e. compare s_i to s_j. If s_i is greater than or equal to s_j, then absorb cluster m_j into cluster m_i, that is for each individual k that belongs to cluster m_j, set $m_k = m_i$. Otherwise, absorb cluster m_i into cluster m_j. Finally, update the vector of cluster sizes, s so that the size of the new cluster equals the sum of the old cluster sizes, $s_i + s_j$.

[4] I refer the interested reader to Newman and Ziff (2001) for details on an even faster algorithm. The algorithm presented in this paper is quite fast and runs in approximately $O(n \log n)$ time; however, Newman and Ziff (2001) show that with some more work it is possible to perform the same calculations in roughly $O(n)$ time.

In the example network depicted in Figure 1.2 panels (a) through (l), the empty network has $m = \{1, 2, 3, 4, 5, 6\}$ and $s = \{1, 1, 1, 1, 1, 1\}$. After the addition of link $(2, 4)$ panel (b), because 2 and 4 do not belong to the same cluster, i.e. $m_2 \neq m_4$, and clusters 2 and 4 are the same size, $s_2 = s_4 = 1$, absorb cluster 4 into cluster 2. The new membership vector is given by $m = \{1, 2, 3, 2, 5, 6\}$, and the new size vector is $s = \{1, 2, 1, \emptyset, 1, 1\}$. After adding link $(1, 2)$, shown in panel (c), because 1 and 2 do not belong to the same cluster, $m_1 \neq m_2$, and cluster 2 is larger than cluster 1, $s_2 > s_1$, absorb cluster 1 into cluster 2. The new membership vector is given by $m = \{2, 2, 3, 2, 5, 6\}$, and the new size vector is $s = \{\emptyset, 3, 1, \emptyset, 1, 1\}$. After adding $(2, 5)$ in panel (d), because 2 and 5 do not belong to the same cluster, $m_2 \neq m_5$, and cluster 2 is larger than cluster 5, $s_2 > s_5$, absorb cluster 5 into cluster 2. The new membership vector is given by $m = \{2, 2, 3, 2, 2, 6\}$, and the new size vector is $s = \{\emptyset, 4, 1, \emptyset, \emptyset, 1\}$, and so on.

With all of the data elements needed to estimate equation (2.1) accounted for, it is possible to measure the bias induced by the unadded (missing) links as links are added. Repeating this process, with new uniformly random sequences of links, traces out the empirical distribution of the bias caused by the missing links for as many values of k as the researcher wishes.

1.5 Empirical application

In this section I apply the algorithm discussed above to study the effects of missing edges on an empirical analysis of US teenagers' recreational and physical activities. The analysis mirrors those of Bramoullé et al. (2009) and Liu et al. (2014) allowing for comparisons across studies. As in these prior studies I use data from the Add Health survey. The survey covers a nationally representative sample of public and private school students in grades 7-12 in the United States. During the 1994-1995 school year students in roughly 130 schools across the country were administered questionnaires on a wide range of topics including behaviors, demographics, family background and friendships.

In the friendship portion of the questionnaire respondents were asked to name their best friends (up to five male and five female) from a school roster. As in Liu et al. (2014), I reconstruct the network of best friends at each school by setting $g_{ij} = 1$ if i nominates j and $g_{ij} = 0$, otherwise. By matching the identification numbers of respondents to their friendship nominations it is possible to obtain information on nominated friends.

The network I observe almost certainly differs from the true network for a number of reasons. As Manski (1993), De Giorgi et al. (2010) and Bramoullé et al. (2020) point out, there is no way to know a priori that the network of best friends is the relevant network. Students might be influenced by any number of individuals including classmates, teachers, siblings, romantic partners and parents.

Moreover, we know that the network of best friends in the Add Heatlh data is measured with error. Roughly 15.7 percent of friendship nominations belong to friends that do not go to a surveyed school, another 11.8 percent belong to friends that go to a surveyed school but either do not appear on a school roster or contain invalid identification numbers. Of the remaining valid nominations less than half are reciprocal. The lack of reciprocity among links could indicate that not all friends influence one another equally or that friendship nominations are unreliable. Either way given that the exclusion restriction for instrumental variables fails when the network is measured with error, one should expect any estimate of equation (2.1) to be biased. The question remains how severely.

To help answer this question I select a sample of the ten schools with the highest percentage of valid nominations. This sample forms the basis for my simulations. After removing clusters of two or fewer students my sample includes 3,061 individuals. My left-hand side variables are indices of recreational and physical activity.

As in Bramoullé et al. (2009) recreational activity is measured by an index of participation in extracurricular activities such as clubs and sports. The index takes values between zero and four, where zero to three corresponds to the number of organizations that a student belongs to and four corresponds to membership in four or more such organizations.[5] Physical activity is measured by the question, "How many times in a normal week do you work, play or exercise hard enough to make you sweat and breath heavily?", coded as never (0), one or two times (1), three to five times (2), six or seven times (3) and more than seven times (4) as in Liu et al. (2014). The right-hand side variables include best friends' mean and aggregate outcomes, own-characteristics: age, gender, race, parents' present at home, parents' level of education and parents' participation in the labor force and best friends' mean characteristics.

Table 3.1 provides descriptive statistics for my sample. The average student is 14.9 years of age with 51 percent identifying as female. Just over 30 percent of students report being a person of color, and 74.9 percent of students report having both parents present at home with comparable levels of labor participation.

Table 2.2 reports two-stage least squares estimates of equation (2.1) with recreational activity as the outcome variable. The local-aggregate and local-average peer effects are both highly statistically significant. The aggregate effect indicates that a one point increase in the sum of best friends' recreational activities leads to a 0.025 increase in own-recreational activities. The average effect induces a 0.468 increase in own-recreational activities for each one point increase in best friends' mean recreational activities. While Bramoullé et al. (2009) do not include aggregate effects in their model, their estimate of the average peer effect is very similar to mine (0.466 vs. 0.468).

[5] An upper bound of four is placed on the number of organizations that a student can belong to in order to minimize the effects of measurement error in the outcome variable (Bramoullé et al., 2009)

Table 2.3 provides two-stage least squares estimates for physical activity. The local-aggregate effect is again, highly statistically significant; however, the average effect is not statistically significant. The endogenous effect indicates a 0.011 increase in own-physical activities for each one point increase in the sum of best friends' physical activity index and a 0.169 increase in own-physical activities for a one point increase in the mean of best friends' physical activity index. My findings match closely with those of Liu et al. (2014) where the aggregate effect was found to be 0.007 and statistically significant at the 1 percent level and the average effect was 0.104 and was not statistically significant.

1.6 Results

By applying my method to the Add Health sample, I am able to estimate the bias caused by links missing at random. I conduct $B = 200$ Monte Carlo simulations for each outcome variable. Table 1.4 reports the mean and standard deviation of the local-aggregate and local-average effect of peers outcomes on own-recreational and physical activities when links are missing at random.

The local-aggregate effect for both recreational and physical activities display a similar pattern. When five percent of the links from the original network are observed the mean estimate is attenuated. For recreational activities the mean estimate is 0.007 or 28 percent of the baseline value. For physical activities the mean estimate is 0.009 or 81 percent of the baseline.

When fifteen percent of links are observed the bias becomes inflationary, reaching its maximum after twenty-five percent of the network is revealed. Thereafter, the bias decreases as links are added to the network. At its maximum, the bias for recreational activities is 0.033 and for physical activities is 0.015 or roughly 30 to 35 percent larger than the baseline value. For recreational activities the bias decreases to about 10 percent once half of the links are observed. The bias for physical activities decreases to about 10 percent after roughly 70 percent of the network is observed. Regardless of the fraction of links observed, the local-aggregate effect is remarkably robust. The mean estimates always lie within one standard deviation of the baseline values, and the standard deviation of the mean estimates decreases exponentially as links are added.

The effect of links missing at random on the local-average effect is more pronounced. Again, both recreational and physical activities exhibit a similar pattern of bias. When five percent of the links from the original network are observed the mean estimates have a strong negative bias. The mean estimate for recreational activities is -0.409 and is -0.402 for physical activities. The mean estimates increase steadily, approaching zero after roughly twenty percent of the network is observed but only reach the baseline value after seventy-five percent of the network is observed.

In addition, the local-average estimates are less precise. At least fifty percent of the links must be observed before the mean estimate for recreational activities is within two standard deviations of the baseline value, and more than seventy-five percent of the links must be observed for physical activities. Furthermore, the standard deviation remains roughly constant around 0.10 until thirty percent or more of the links are observed, and falls more slowly than that of the local-aggregate effect.

Given these findings and the error rates observed in the Add Health data, one could expect that the local-aggregate effects reported in Tables 2.2 and 2.3 and by Liu et al. (2014) are close to the true values. One might also expect that the local-average effects reported in Table 2.2 and by Bramoullé et al. (2009) are on the order of the true values, but the variance is likely understated.

1.6.1 Robustness

In order to determine whether the relationship between the number of links missing at random and estimates of the local-aggregate and local-average effect hold more broadly, I simulate network data and repeat the above experiment with different values for λ_1 and λ_2. I begin by constructing Erdős and Rényi random graphs containing $n = 1000$ individuals and $\ell = 4,000$ links divided evenly among five clusters with links placed uniformly at random. I then set the following parameters to be held constant throughout the simulations: $\beta = 0.25$, $\gamma = 0.15$, $\delta = 1$ and $\sigma^2 = 0.009$ with $X \sim N(1, 3)$.

In the first set of simulations I calculate the "true values of Y for $\lambda_1 = \{0.01, 0.05, 0.10\}$ and $\lambda_2 = 0.5$. I then apply my method and conduct $B = 500$ simulations for each value of λ_1. The results are reported in Table 1.5.

Again a pattern of initial inflation emerges. The simulated data reaches its maximum after only ten percent of the network is observed. The magnitude of the bias is inversely proportional to the size of λ_1. When $\lambda_1 = 0.01$ the maximum mean estimate is 0.040 or an increase of roughly 300 percent. When $\lambda_1 = 0.1$ the maximum mean estimate is 0.111 or an increase of 11 percent. The inflationary bias shrinks as more links are observed and disappears once fifty percent of the links are observed. When more than fifty percent of the links are observed the initial inflationary bias may be replaced by a small attenuation bias. The magnitude of the downward bias is again inversely proportional to the value of λ_1. When $\lambda_1 = 0.01$ the mean estimates may be as much 40 percent lower than the baseline values. When $\lambda_1 = 0.1$ the mean estimates never decrease by more than 1 percent of the baseline value. The simulated local-aggregate effects are not as precise as the Add Health estimates, but they are still precise enough that regardless of the percentage of links observed the mean estimate is always within two standard deviations of the baseline value.

To test the robustness of the local-average effects, I repeat the above experiment with $\lambda_1 = 0.05$ and let $\lambda_2 = \{0.1, 0.3, 0.5\}$ vary. As in the Add Health data, the local-average effects are far less precise than the aggregate effects. The mean estimate is only within two standard deviations of the observed value when ninety percent or more of the links are observed.

The negative bias that was present in the Add Health data is less severe in the simulated network data. The bias is largest when $\lambda_2 = 0.1$ and decreases in magnitude as λ_2 increases; however, the bias also appears to be more persistent. When $\lambda_1 = 0.1$ the minimum mean estimate is -0.086 when 25 percent of the links are observed and remains negative until seventy percent of the links are observed. When $\lambda_2 = 0.3$ the mean estimates take on positive values after fifty percent of the links are observed and are always positive when $\lambda_2 = 0.5$. Compared to the Add Health data much more of the Erdős and Rényi networks must be observed in order to obtain a positive mean estimate.

1.7 Discussion

This study explores a new method for estimating the bias caused by measurement error in a generalized peer effects model. My approach recognizes that adding a new link to an existing network is analogous to a step in the execution of an efficient sparse matrix multiplication algorithm. By adding links one at a time I can enumerate the changes in the vectors of peers outcomes and characteristics, and by adding links in the "right" order I can estimate the effects of links missing at random.

To my knowledge, this is the first study in the peer effects literature to estimate the effect of links missing at random.[6] I apply my method to data from the Add

[6] De Giorgi et al. (2010) estimate the effects of measurement error when links may be missing or spurious at random.

Health survey and simulated network data. I find that both the local-aggregate and local-average effects exhibit clear patterns of bias. The local-aggregate effect has a modest inflationary bias when only a fraction of the total number of links are observable. The bias reaches its maximum when the observed network contains between 10 and 25 percent of the total number of links. The bias decreases when more links are observed and vanishes once 50 percent or more of the links are revealed.

The local-average effect suffers from a persistent negative bias. At least 75 percent of the total links in the Add Health data and 90 percent of the links in the random graphs must be observed to eliminate the downward bias. The local-average effects also suffer from far more variance than the local-aggregate estimates.

My findings are good news for researchers who expect to observe 90 percent or more of the links in a given network or who expect to observe at least half of the links in a network and are only interested in estimating local-aggregate effects. In either of these cases the effect of links missing at random should be negligible. When less than half of the links in a given network are observed researchers can expect local-aggregate effects to be overstated by as much as 30 percent and local-average effects to be understated and imprecisely estimated. Researchers are encouraged to estimate the bias in their own data. Sample code used to construct tables 1.5 and 1.6 are available on my GitHub page.

My findings add to the relative dearth of literature on measurement error in peer effects models. While this study provides further evidence for the relative robustness of local-aggregate models to measurement error, peer effect models contain many moving parts. More research is needed to determine if local-aggregate effects are robust to other forms of measurement error and the degree to which local-average effects may be biased.

Table 1.1
Descriptive statistics

Variable	Mean	Std. Dev.
Recreational activities	2.309	1.458
Physical activity	2.373	1.191
Age	14.934	1.743
Female	0.510	0.500
White only	0.698	0.459
Mom and Dad present at home	0.749	0.434
Mom has a college degree	0.220	0.414
Dad has a college degree	0.199	0.399
Mom works	0.737	0.441
Dad works	0.724	0.447
Number of observations 3,061		

Data was obtained from a sample of ten schools with the lowest reported error rates among friendship nominations drawn from the Add Health Wave 1 In School questionnaire. Recreational activity is measured by an index of participation in extracurricular activities such as clubs and sports with an upper limit of four as in Bramoullé et al. (2009). Physical activity is measured by the question, "How many times in a normal week do you work, play or exercise hard enough to make you sweat and breath heavily?", coded as never (0), one or two times (1), three to five times (2), six or seven times (3) and more than seven times (4) as in Liu et al. (2014). The unit of observation is an individual.

Table 1.2
Estimation of peer effects in recreational activities

	Variable	Coefficient	SE
Local-aggregate effect (λ_1)	Recreational activities	0.025***	0.004
Local-average effect (λ_2)	Recreational activities	0.468***	0.092
Own-characteristics (β)	Age	0.010	0.017
	Female	0.280***	0.051
	Race is white only	-0.170**	0.078
	Mom and dad present	-0.024	0.093
	Mom has college degree	0.214***	0.063
	Dad has college degree	0.250***	0.066
	Mom works	0.103*	0.057
	Dad works	0.124	0.089
Average neighbors' characteristics (γ)	Age	-0.066***	0.012
	Female	-0.117	0.106
	Race is white only	-0.171	0.121
	Mom and dad present	0.006	0.194
	Mom has college degree	0.210	0.140
	Dad has college degree	0.393***	0.152
	Mom works	0.028	0.123
	Dad works	0.002	0.182

Effect of peer recreational activity on own-activity. The unit of observation is an individual. Data was obtained from a sample of ten schools with the lowest reported error rates among friendship nominations drawn from the Add Health Wave 1 In School questionnaire. Recreational activity is measured by an index of participation in extracurricular activities such as clubs and sports with an upper limit of four as in Bramoullé et al. (2009). Standard errors are heteroscedasticity robust. Statistical significance: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 1.3
Estimation of peer effects in physical activities

	Variable	Coefficient	SE
Local-aggregate effect (λ_1)	Physical activities	0.011***	0.003
Local-average effect (λ_2)	Physical activities	0.169	0.109
Own-characteristics (β)	Age	-0.049**	0.021
	Female	-0.709***	0.043
	White only	0.058	0.065
	Mom and dad present	0.106	0.078
	Mom college degree	0.133**	0.052
	Dad college degree	-0.002	0.055
	Mom works	-0.025	0.047
	Dad works	-0.006	0.075
Average neighbors' characteristics (γ)	Age	-0.037**	0.015
	Female	0.144	0.103
	White only	-0.144	0.112
	Mom and dad present	-0.075	0.163
	Mom college degree	0.125	0.112
	Dad college degree	0.060	0.116
	Mom works	-0.052	0.099
	Dad works	0.160	0.153

Effect of peer physical activity on own-activity. The unit of observation is an individual. Data was obtained from a sample of ten schools with the lowest reported error rates among friendship nominations drawn from the Add Health Wave 1 In School questionnaire. Physical activity is measured by the question, "How many times in a normal week do you work, play or exercise hard enough to make you sweat and breath heavily?", coded as never (0), one or two times (1), three to five times (2), six or seven times (3) and more than seven times (4) as in Liu et al. (2014). Standard errors are heteroscedasticity robust. Statistical significance: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 1.4
Estimation of peer effects with links missing at random

	Recreational activities					Physical activities			
% of	$[\lambda_1]$		$[\lambda_2]$		% of	$[\lambda_1]$		$[\lambda_2]$	
links	Mean	SD	Mean	SD	links	Mean	SD	Mean	SD
5	0.007	0.050	-0.409	0.104	5	0.009	0.040	-0.402	0.095
10	0.017	0.026	-0.294	0.089	10	0.008	0.023	-0.250	0.089
15	0.030	0.015	-0.145	0.100	15	0.013	0.014	-0.093	0.095
20	0.032	0.011	0.004	0.101	20	0.014	0.011	-0.018	0.092
25	0.033	0.010	0.108	0.100	25	0.015	0.009	0.026	0.100
30	0.032	0.009	0.216	0.105	30	0.014	0.007	0.063	0.104
35	0.030	0.008	0.279	0.101	35	0.014	0.007	0.090	0.095
40	0.029	0.006	0.344	0.093	40	0.014	0.005	0.108	0.086
45	0.028	0.005	0.374	0.084	45	0.014	0.004	0.125	0.082
50	0.028	0.005	0.402	0.085	50	0.014	0.004	0.139	0.072
55	0.027	0.004	0.433	0.070	55	0.013	0.003	0.155	0.070
60	0.027	0.003	0.448	0.069	60	0.013	0.003	0.155	0.071
65	0.027	0.003	0.459	0.079	65	0.013	0.003	0.153	0.069
70	0.027	0.003	0.462	0.068	70	0.013	0.003	0.154	0.066
75	0.026	0.002	0.473	0.064	75	0.012	0.002	0.160	0.061
80	0.026	0.002	0.476	0.052	80	0.012	0.002	0.168	0.054
85	0.025	0.002	0.479	0.045	85	0.012	0.001	0.166	0.048
90	0.025	0.002	0.481	0.042	90	0.011	0.001	0.170	0.038
95	0.025	0.001	0.474	0.034	95	0.011	0.001	0.166	0.030
100	0.025	-	0.468	-	100	0.011	-	0.169	-

Monte Carlo simulations of the effect of peer recreational and physical activity on own-activity when links are missing at random. Data was obtained from a sample of ten schools with the lowest reported error rates among friendship nominations drawn from the Add Health Wave 1 In School questionnaire. Recreational activity is measured by an index of participation in extracurricular activities such as clubs and sports with an upper limit of four as in Bramoullé et al. (2009). Physical activity is measured by the question, "How many times in a normal week do you work, play or exercise hard enough to make you sweat and breath heavily?", coded as never (0), one or two times (1), three to five times (2), six or seven times (3) and more than seven times (4) as in Liu et al. (2014). The unit of observation is an individual, $n = 3,061$ with $\ell = 16,777$ links. $B = 200$ repetitions were performed for each outcome.

Table 1.5
Network simulations with links missing at random

% of links	[$\lambda_1 = 0.01$] Mean	SD	[$\lambda_1 = 0.05$] Mean	SD	[$\lambda_1 = 0.10$] Mean	SD
5	0.031	0.092	0.053	0.105	0.079	0.164
10	0.040	0.042	0.071	0.049	0.111	0.081
15	0.034	0.023	0.066	0.027	0.109	0.047
20	0.030	0.016	0.064	0.018	0.107	0.032
25	0.026	0.012	0.061	0.014	0.105	0.024
30	0.021	0.010	0.058	0.011	0.105	0.019
35	0.018	0.008	0.055	0.009	0.103	0.015
40	0.014	0.007	0.052	0.007	0.102	0.012
45	0.011	0.006	0.050	0.006	0.101	0.010
50	0.009	0.005	0.049	0.005	0.100	0.008
55	0.007	0.004	0.048	0.004	0.100	0.007
60	0.006	0.004	0.047	0.003	0.100	0.006
65	0.006	0.003	0.047	0.003	0.099	0.005
70	0.006	0.003	0.047	0.003	0.099	0.004
75	0.006	0.002	0.047	0.002	0.099	0.003
80	0.007	0.002	0.047	0.002	0.099	0.003
85	0.008	0.002	0.048	0.002	0.099	0.002
90	0.009	0.002	0.049	0.001	0.099	0.002
95	0.009	0.001	0.049	0.001	0.100	0.001
100	0.010	-	0.050	-	0.100	-

Monte Carlo simulations of the effect of peer outcomes on own-outcomes with links missing at random. Each network contains five Erdős and Rényi random graphs with $n = 200$ individuals and $\ell = 800$ links placed uniformly at random. Mean and standard deviations based on $B = 500$ repetitions with $\lambda_2 = 0.5$, $\beta = 0.25$, $\gamma = 0.15$, $\delta = 1$ and $\sigma^2 = 0.009$ held constant while varying λ_1.

Table 1.6
Network simulations with links missing at random continued

% of links	$[\lambda_2 = 0.1]$ Mean	SD	$[\lambda_2 = 0.3]$ Mean	SD	$[\lambda_2 = 0.5]$ Mean	SD
5	-0.043	0.168	-0.028	0.170	0.051	0.272
10	-0.073	0.087	-0.058	0.092	0.006	0.150
15	-0.078	0.058	-0.057	0.060	0.010	0.105
20	-0.085	0.048	-0.060	0.050	0.015	0.085
25	-0.086	0.042	-0.054	0.043	0.025	0.074
30	-0.086	0.041	-0.047	0.042	0.030	0.071
35	-0.083	0.040	-0.036	0.041	0.044	0.066
40	-0.075	0.040	-0.020	0.041	0.056	0.065
45	-0.065	0.038	-0.001	0.039	0.072	0.065
50	-0.052	0.037	0.021	0.038	0.093	0.063
55	-0.037	0.037	0.046	0.037	0.117	0.062
60	-0.022	0.036	0.071	0.037	0.143	0.061
65	-0.005	0.035	0.100	0.036	0.176	0.061
70	0.014	0.035	0.130	0.035	0.216	0.061
75	0.032	0.036	0.161	0.036	0.261	0.061
80	0.052	0.033	0.194	0.033	0.308	0.060
85	0.069	0.030	0.224	0.030	0.356	0.057
90	0.085	0.025	0.254	0.026	0.407	0.052
95	0.100	0.018	0.284	0.019	0.461	0.038
100	0.112	-	0.31	-	0.505	-

Monte Carlo simulations of the effect of peer outcomes on own-outcomes with links missing at random. Each network contains five Erdős and Rényi random graphs with $n = 200$ individuals and $\ell = 800$ links placed uniformly at random. Mean and standard deviations based on $B = 500$ repetitions with $\lambda_1 = 0.05$, $\beta = 0.25$, $\gamma = 0.15$, $\delta = 1$ and $\sigma^2 = 0.009$ held constant while varying λ_2.

CHAPTER 2

NON-RANDOM ERRORS IN PEER EFFECTS

2.1 Introduction

Decades of research has sought to understand the relationship between individual and peer behavior. Much of this research has grappled with problems identifying peer effects (Blume et al., 2015). Unobserved common shocks, endogenous group formation and the reflection problem, the inability to distinguish between the influence of peers outcomes and peers characteristics when observing group averages, all pose serious threats to identification (Manski, 1993). Some of these issues can be circumvented when more detailed knowledge of the network structure is available (Bramoullé et al., 2009; Liu and Lee, 2010; De Giorgi et al., 2010; Laschever, 2011). A small but growing literature has raised concerns that misspecifying the network structure violates the identifying assumptions of these models (De Giorgi et al., 2010; Chandrasekhar and Lewis, 2011; Bramoullé et al., 2020). Moreover, there is reason to believe that misspecifications are a ubiquitous problem in applied research that deserves further investigation (De Giorgi et al., 2010; Bramoullé et al., 2020).

Misspecification can occur for a number of reasons including the sampling method, top-coding the number of links, network boundary constraints, mis-coding or mis-reporting links and non-response (Advani and Malde, 2018). The first three issues are related to the sampling design and have been discussed in Chandrasekhar and Lewis (2011), Liu (2013), Patacchini et al. (2017), Advani and Malde (2018) and Griffith (2019). By contrast there is a dearth of research on mis-coding and mis-reporting errors. This paper and Chapter 1 seek to help fill this hole in the literature.

Chapter 1 explores a novel method for generating the counterfactual networks needed to estimate the bias caused by links missing at random. In this paper I show how the techniques introduced in Chapter 1 can be extended to study other forms of measurement error. I use the problem of top-coding the number of reported links to motivate the methodology. I then extend the technique to study missing links when mis-reporting is correlated with the network structure. Finally, I consider the impact of mis-reporting that results from the presence of spurious links.

Throughout the paper I employ data from the National Longitudinal Study of Adolescent to Adult Health (Add Health) and a series of simulated networks to empirically estimate the bias. I present my findings via a generalized peer effects model due to Liu et al. (2014). The model nests three other work-horse models, the linear-in-means model of Manski (1993), the local-aggregate model of Liu and Lee (2010) and the local-average model of Bramoullé et al. (2009). The choice of model serves to ensure that my results are generalizable to as much of the literature as possible.

My paper contributes to the measurement error in peer effects literature in three ways. First, I provide empirical evidence on the impact of measurement error in previously understudied sources of misspecification. Second, I demonstrate the flexibility of the methods introduced in Chapter 1 by extending those techniques to three new applications. Third, my findings generalize insights gleaned from the prior literature to a broader class of peer effects models.

The rest of the paper is organized as follows. Section 2 describes the notation and empirical model used throughout the rest of the paper. Section 3 briefly reviews the Methods used in Chapter 1. Section 4 introduces the empirical application of the methods outlined in Section 3. Section 5 introduces the concept of top-coding and provides evidence of its effect on the empirical model. Section 6 discusses a stylized example of mis-reporting that is correlated with the network structure and its effects on the empirical model. Section 7 extends the methods employed thus far to study the effects of spurious links, and Section 8 concludes.

2.2 Empirical framework

In this section I introduce the notation and empirical model. Throughout the paper I restrict my attention to networks that consist of a set of n individuals whose interactions are by assumption binary. The set of interactions may then be represented by either a $n \times n$ adjacency matrix, G, with typical element $g_{ij} = 1$ if i is linked to j and 0 otherwise or by a list, ℓ, of ordered node pairs with typical element (i, j) indicating that i is linked to j. Note that as in Liu et al. (2014) I do not require links to be symmetric. Although the adjacency matrix form is more common in the peer effects literature and allows one to express such models using compact notation, the list form ℓ will be instrumental to the methods discussed in Section 3.[1] See Figure 2.1 for examples of each form.

The set of individuals that i is linked to is called their neighborhood, denoted $n(i)$. Let g_i equal the size of neighborhood i, $|n(i)|$.[2] Finally, given the adjacency matrix, G, and the vector $g = (g_1, g_2, \ldots, g_n)'$, the row-normalized adjacency matrix A is defined such that each element $a_{ij} = g_{ij}/g_i$.

With these definitions one can write a generalized peer effects model as in Liu et al. (2014)

$$Y = \lambda_1 GY + \lambda_2 AY + X\beta + AX\gamma + \delta + \epsilon \quad , \tag{2.1}$$

where Y is a vector of outcomes, G and A are as defined above, so that GY is a vector of the sums of neighbors' outcomes with i^{th} element equal to $\sum_{j \in n(i)} y_j$, AY is a vector of the averages of neighbors' outcomes with i^{th} element equal to $g_i^{-1} \sum_{j \in n(i)} y_j$, X is a matrix of exogenous characteristics, AX is a matrix of the averages of neighbors' characteristics with i^{th} element equal to $g_i^{-1} \sum_{j \in n(i)} x_j$, δ is a cluster fixed effect and ϵ is an i.i.d. vector of innovations.[3]

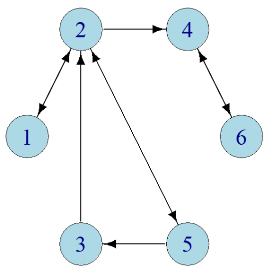
The parameters λ_1, λ_2 and γ capture peers' influence. When λ_1 is non-zero it is called a local-aggregate effect because the effect is a function of neighbors aggregate outcomes. The presence of local-aggregate effects suggest that the network topology plays an important role in determining behavior, and individual policy shocks can affect the entire network. When λ_2 is non-zero it is referred to as a local-average effect because the effect depends on neighbors average outcomes. Local-average effects can indicate the presence of social norms that allow network level shocks to persist while dampening individual shocks. When γ is non-zero

[1] Note that I do not require interactions to be symmetric; however, it is straightforward to apply the ideas in this paper to symmetric networks.

[2] In the networks literature g_i is commonly referred to as the degree of i if G is symmetric or the out-degree of i if G is asymmetric.

[3] See Liu et al. (2014) for a treatment of the micro-foundations of the model in equation (2.1).

Figure 2.1
A simple network depicted as a graph, an adjacency matrix and a list



$$G = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

$$\ell = \begin{array}{c} \boxed{(1,\ 2)} \\ \boxed{(2,\ 1)} \\ \boxed{(2,\ 4)} \\ \boxed{(2,\ 5)} \\ \boxed{(3,\ 2)} \\ \boxed{(4,\ 6)} \\ \boxed{(5,\ 2)} \\ \boxed{(5,\ 3)} \\ \boxed{(6,\ 4)} \end{array}$$

A simple network with six nodes and nine edges represented as a graph (left panel), a 6×6 adjacency matrix, G, (center panel) and a nine element list, ℓ, (right panel).

the model is said to exhibit exogenous peer effects, where individual behavior is influenced by the exogenous characteristics of their peer group.

2.2.1 Identification

The peer effects in equation (2.1) cannot usually be estimated via OLS because of the simultaneous behavior of peers (Manski, 1993). Instead, Liu et al. (2014) propose an instrumental variables approach. Identification depends on detailed knowledge of and sufficient variation in the network structure. The key idea is to use neighbors of neighbors and variation in the size of neighborhoods as instruments.

For the purpose of illustration, suppose that Ian and Jon are friends with Kat, but not friends with one another. Then Ian and Jon have no direct influence on one another, an exclusion restriction, but can influence one another indirectly via Kat, i.e. they are relevant to one another. The indirect nature of Ian and Jon's relationship forms a set of potential instruments for the local-average effect, while the variation in neighborhood sizes forms a set of potential instruments for the local-aggregate effect.

Formally, Liu et al. (2014) prove that when the following conditions hold, the parameters in equation (2.1) are identified. Condition 1 states that the matrices I, G, A, GA, A^2, GA^2 and A^3 are linearly independent. When Condition 1 holds their is sufficient variation in the neighborhoods of interconnected peers to satisfy the exclusion restriction. Condition 2 states that either $\lambda_1\beta \neq 0$ or $\lambda_2\beta + \delta \neq 0$. When Condition 2 holds a set of relevant predictors exists.

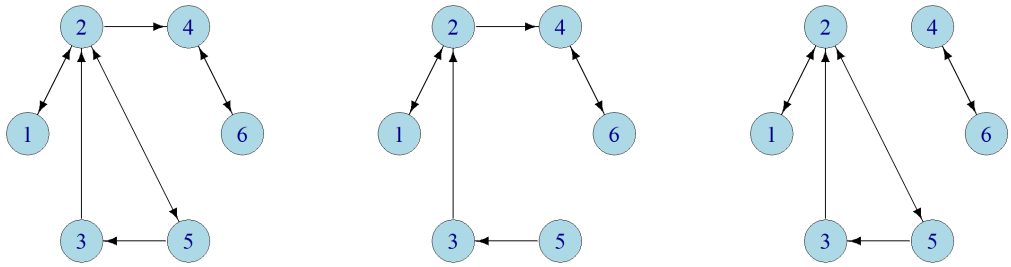
When Conditions 1 and 2 are met the vector of degrees, $g = (g_1, g_2, \ldots, g_n)'$, and the sum of neighbors characteristics, captured in GX, can be used to instrument for the sum of neighbors outcomes, GY, and the average of neighbors' average neighbors' characteristics, A^2X, can be used to instrument for the average of neighbors' outcomes, AY.

2.2.2 Measurement Error

Implicit in the instrumental variables approach is the assumption that the network is observed without error. When the network is measured with error the exclusion and relevance conditions outlined above may be violated leading to biased estimates of equation (2.1) (Blume et al., 2015). To illustrate the problem posed by measurement error consider the graphs in Figure 2.2. Figure 2.2 panel (a) contains the example network from Figure 2.1 redrawn for reference. Panels (b) and (c) illustrate possible misspecifications of the example network in panel (a). In panel (b) the missing links between individuals 2 and 5 violate the exclusion restriction because 2 is directly connected to 5 in the original graph resulting in biased instruments.[4] In panel (c) the missing link between individuals 2 and 4 eliminates relevant predictors leading to less informative instruments. The relative frequency of creating invalid instruments and of destroying valid instruments will depend on

[4]Only the (5,2) link needs to be missing to violate the exclusion restriction. However, in order to make it easier to see the difference between panels (a) and (b) both the (5,2) and (2,5) links are missing.

Figure 2.2
A simple network



(a) (b) (c)

Panel (a) recalls the example network from Figure 2.1. Panels (b) and (c) depict two possible misspecifications of the network in Panel (a). Panel (b) illustrates a violation of the exclusion restriction between individuals 2 and 5. Panel (c) represents a misspecification that leads to less informative instruments due to the loss of relevant predictors.

the network topology and the source of misspecification. Although no analytic solution to address such errors exists, the bias can be estimated empirically.

2.3 Methods

De Giorgi et al. (2010), Liu (2013) and Griffith (2019) estimate the empirical distribution of the bias caused by measurement error via Monte Carlo simulations. They assume that Y, X and the true network structure are known. They then define a stochastic process that perturbs the true network structure in a way that mimics the form of measurement error of interest.

Chapter 1 builds on their work by proposing a method for efficiently generating networks that contain links missing at random. I briefly review the main insights of Chapter 1 here. In later sections I demonstrate how these ideas can be extended to study other forms of measurement error.

The first insight from Chapter 1 is that the addition of a link (i, k) to a given network only changes the values in row i of equation (2.1). For example, row i of GY, is equal to the sum of i's neighbors outcomes, i.e. $\sum_{j \in n(i)} y_j$. Adding link (i, k) adds y_k to this sum. The other $n-1$ rows are unaffected because only row i depends on the neighborhood of i. Similar arguments follow for AY, AX and the instruments g and GX. Chapter 1 goes on to show that the number of impacted rows in the instrument A^2X is proportional to g_i, and that by applying methods developed in Newman and Ziff (2001) the cluster fixed effects can also be tracked in roughly constant time. As a result, given Y, X and an ordered list of interactions, ℓ, one can efficiently enumerate the changes in the regressors as the links in ℓ are placed on the network one at a time.

The second insight from Chapter 1 deals with the relation between the order of ℓ and the interpretation of estimates based on equation (2.1) when only k links are observed. When ℓ is ordered uniformly at random and k links are observed, the set of $\ell - k$ missing links are selected as if they were drawn uniformly at random. The difference between estimates of equation (2.1) based on the original network and the network with $\ell - k$ missing links measures the bias induced by those missing links. Repeated estimates, each time with a new uniformly random order for ℓ allows one to characterize the empirical distribution of the bias. Furthermore, because adding links one at a time enumerates the changes in the regressors the empirical distribution of the bias can be estimated for multiple values of k simultaneously.

2.4 Empirical Application

The rest of the paper is devoted to extending the methods outlined above to study other forms of measurement error. I begin by showing how links can be ordered to estimate the effects of top-coding. I then examine the effects of missing links when the likelihood of mis-reporting is non-random. Finally, I explore the effects of spurious links.

Throughout the paper I estimate the resulting bias on US teenagers' recreational and physical activities. I use data from the Add Health survey to catalog my results. The Add Health survey is one of the most commonly cited data sources in the peer effects literature. The survey includes a nationally representative sample of middle and high school students at public and private schools across the US during the 1994-1995 school year. Survey topics include questions on behavior, demographics, family and friends.

The friendship questionnaire asks students to name their five best male and five best female friends.[5] As in Liu et al. (2014) and Chapter 1 I reconstruct the network of best friends based on these responses. I set $g_{ij} = 1$ if student i nominated student j and $g_{ij} = 0$, otherwise.

For ease of comparison, I use the same sample of students and the same left-hand side variables as in Chapter 1. The sample includes the 10 schools with the highest percentage of valid friendship nominations and a total of 3,061 students.[6] The left-hand side variables are indices of recreational and physical activity. I define recreational activity as in Bramoullé et al. (2009) as an index of participation in extracurricular activities such as clubs and sports. The index takes values between zero and four, where zero to three correspond to the number of organizations that a student belongs to and four corresponds to membership in four or more such organizations.[7] Physical activity is measured as in Liu et al. (2014) by the question, "How many times in a normal week do you work, play or exercise hard enough to make you sweat and breath heavily?", coded as never (0), one or two times (1), three to five times (2), six or seven times (3) and more than seven times (4). My right-hand side variables include best friends' aggregate

[5]Responses were limited to a maximum of five male and five female best friends.

[6]Friendship nominations may be invalid because the nominated friend does not go to a surveyed school, the friend goes to a surveyed school but is not on the Add Health roster or because the identification number used is out-of-range.

[7]Bramoullé et al. (2009) place an upper bound of four to minimize the effects of measurement error in the outcome variable

and mean outcomes, own-characteristics: age, gender, race, parents' present at home, parents' level of education and parents' participation in the labor force and friends' mean characteristics.

Descriptive statistics are reported in Table 3.1. The average student is 14.9 years of age with 51 percent identifying as female. Just over 30 percent of students report being a person of color, and 74.9 percent of students report having both parents present at home with comparable levels of labor participation.

Tables 2.2 and 2.3 report two-stage least squares estimates of equation (2.1) with recreational and physical activity as the outcome variables, respectively. While the local-aggregate effect for both recreational and physical activity is highly statistically significant, only the local-average effect of recreational activities is significant. The point estimates of λ_1 and λ_2 form a baseline for comparing the bias caused by measurement error throughout the rest of the paper.

2.5 Top-coding neighborhood size

In this section I adapt the methods in Chapter 1 to study the top-coding problem introduced in Griffith (2019). Griffith highlights how researchers often rely on survey responses to reveal the underlying network structure. Surveys by nature are not exhaustive, and as a result, frequently lead to various forms of measurement error (Chandrasekhar and Lewis, 2011; Liu, 2013; Griffith, 2019). For example, many surveys limit the number of peers that respondents can report. The Add Health survey limits respondents to name up to five male and five female friends. Other surveys used in the literature impose similar limitations. For example, Banerjee et al. (2013) allow up to eight peers depending on the context. Oster and Thornton (2012) limit responses to a maximum of three close friends. While, Cai et al. (2015) and Kandpal and Baylis (2013) permit respondents to list up to five friends. When these censoring rules are binding, they can bias parameter estimates.

Griffith (2019) shows that in the Add Health data a majority of students list the maximum number of friends suggesting that the Add Health limit is binding. Although the network that permits students to report more than five male and five female friends is unobservable, one can extrapolate the potential bias by enforcing stricter limitations on the observable data. For example, one can construct the network where students only report four male and four female friends from the observable data and compare those results with the results from the unaltered network.

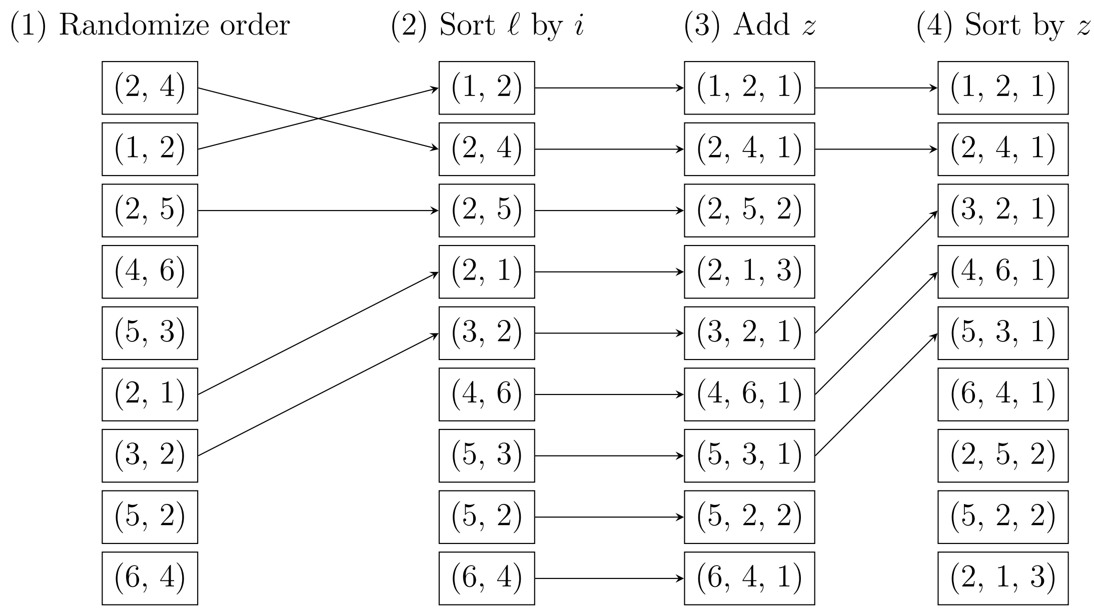
2.5.1 Ordering links to mimic top-coding

I now propose a method for ordering the links in ℓ that when combined with the techniques from Chapter 1 can estimate the bias caused by top-coding. The goal is to order ℓ such that one peer is added to every neighborhood of size $g_i \geq 1$ before a second peer is added to any other neighborhood. Similarly, a second peer is added to each neighborhood of size $g_i \geq 2$ before a third peer is added to any neighborhood, and so on. When the order in which neighbors are reported is uncorrelated with their characteristics, what Griffith calls order irrelevance, then the following sorting method produces the desired result.

First, sort ℓ uniformly at random. Next, sort ℓ by neighborhood, that is by the first element, i, in each node pair, maintaining the uniformly random order within neighborhoods. Maintaining the random order within neighborhoods ensures order irrelevance. Within each neighborhood assign each link a number from one to g_i. Call this number z. The number z indicates whether the link corresponds to the first link reported, the second link, and so on. Finally, sort ℓ by z maintaining the order within values of z. See Figure 2.3 for an illustration of the sorting method.

Forming the network with only the links in ℓ that satisfy $z = 1$ mimics the effect of limiting respondents to report only one neighbor. An estimate of equation (2.1) based on this network measures the potential bias. The bias when the network contains only links that satisfy $z \leq 2$ estimates the effect of limiting respondents to two neighbors, and so on. Repeating the process many times allows one to estimate the empirical distribution of the bias for limits from one to $\max_i g_i$.

Figure 2.3
Example sorting ℓ to mimic censoring neighborhood size



In step (1) ℓ is sorted uniformly at random. In step (2) ℓ is sorted by the first element, i, in each node pair. In step (3) the counter, z, is added to each node pair, and finally, in step (4) ℓ is sorted by z. Although all elements are sorted in each step, I only display arrows linking the first five elements in steps (2) and (4) for readability.

2.5.2 Application to Add Health

The Add Health data allows respondents to report up to five male and five female friends. To ensure that the network admits the appropriate number of male and female friends, I apply the sorting method above to two lists, a list of male friends, ℓ_m, and a list of female friends, ℓ_f. After assigning each link a value z in Step (3) above I combine the lists to form $\ell = (\ell_m', \ell_f')'$ before sorting ℓ on z in Step (4).

The resulting list admits up to one male and one female friend when $z = 1$, up to two male and two female friends when $z = 2$ and so on. I conduct 200 Monte Carlo simulations applying this method to the Add Health data and report my findings in Table 2.4.

Similar to Griffith (2019)'s finding that the effect of neighbors' average characteristics, γ, tend to be downward biased, I find that when respondents are restricted to report fewer than five male and female friends the local-average effect, λ_2, also trends toward zero. Estimates of the local-average effect are 30 to 70 percent smaller when only one male and one female friend are reported compared to the baseline estimates. As more friends are permitted the bias decreases and once three or more friends are reported the estimates are within two standard deviations of the baseline estimates.

Unlike the exogenous and local-average effects, the local-aggregate effect, λ_1, is overestimated when fewer than five male and five female friends are reported. When only one male and one female friend are reported the estimated effect is more than four times the baseline value for recreational activities and more than double the baseline value for physical activities. After two male and two female friends are permitted the bias falls to 88 and 55 percent respectively and then to 36 and 18 percent when three male and three female friends are reported. The sensitivity of the local-aggregate effect to top-coding stands in stark contrast to the findings of Griffith (2019) and Chapter 1, where the linear-in-means model almost always underestimates the baseline effect and links missing at random have a relatively mild inflationary bias.

2.5.3 Robustness checks

As in Chapter 1, I also estimate the effects of top-coding on a set of Erdős-Rényi random graphs. These simulated networks serve as a robustness check on the

Add Health data and are intended to provide readers with additional confidence about the generality of my findings. The simulated networks are constructed with $n = 1,000$ individuals and $\ell = 4,000$ links placed uniformly at random across five equally sized clusters. Throughout the network simulations I hold the following parameters constant: $\beta = 0.25$, $\gamma = 0.15$, $\delta = 1$ and $\sigma^2 = 0.009$ with $X \sim N(1, 3)$, while varying λ_1 and λ_2.

Mean and standard deviations for the local-aggregate effect with λ_1 equal to 0.01, 0.05 or 0.1 with λ_2 held constant at 0.5 are reported in Table 2.5. When individuals are limited to reporting only one friend the identifying assumptions from Section 2.1 completely break down. The instruments and regressors become almost perfectly collinear leading to a form of the reflection problem and severe inflation of the local-aggregate effect.[8] Once two or more friends are reported the bias continues to overestimate the true effect but is less severe. When up to six friends may be reported the bias falls to a less than 20 percent increase in the estimate. As more friends are reported the bias decreases approaching the true value from above.

Mean and standard deviations for the local-average effect with λ_2 equal to 0.1, 0.3 or 0.5 and λ_1 held constant at 0.05 are reported in Table 2.6. When individuals are limited to reporting only one friend A^2X is not of full rank and the local-average effect is not estimable. When individuals are limited to reporting only two friends and λ_2 is small the local-average effect is estimable but is overestimated. With three or more friends the bias is attenuating and increasing in the number of reported friends. Aside from the failure of the identification conditions when only one friend is reported, the bias in the simulated networks is qualitatively similar to that of the Add Health data.

[8]Note that this does not occur in the Add Health data because individuals are permitted to report one male and one female friend.

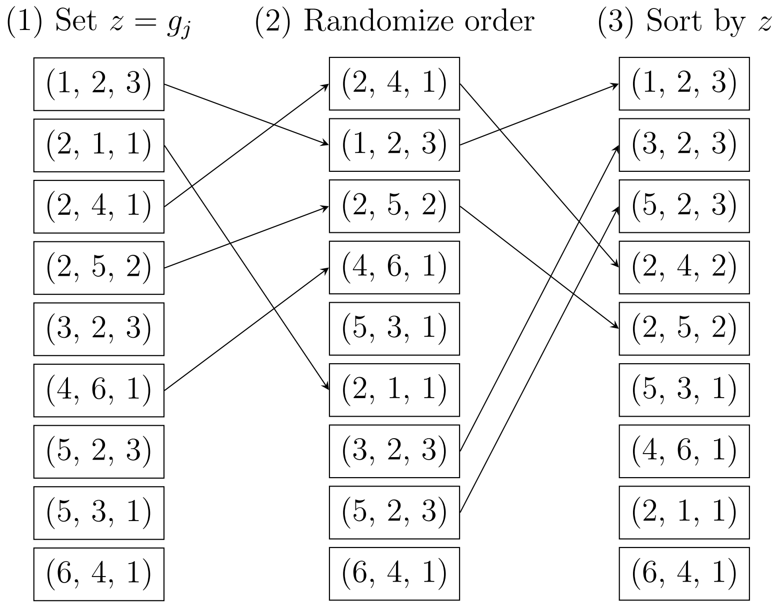
2.6 Links missing not at random

In the real world links may not be missing at random. In order to gain insight into the effects of missing links when mis-reporting is not random I consider the following stylized scenario. I imagine that individuals might prioritize reporting links that they deem important. The notion of importance that I have in mind is one where the more neighbors a person has, the more important they are and by extension the more important links pointing to them are. One might expect reporting to be positively correlated with neighborhood size in this way when individuals care about how well connected they appear.

In order to make the model tractable, I assume that there is a fixed fraction of missing links and that the likelihood of reporting a link is directly proportional to the neighborhood size of the linked individual. To order links in this way, first assign each link in ℓ a number z that equals the neighborhood size of the *second* individual in each node pair, that is z equals g_j. Next sort ℓ uniformly at random. Finally, sort ℓ by z in descending order maintaining the random order within equal values of z. Ordering by z generates the desired correlation between neighborhood size and the likelihood of a link being observed by ensuring that links that point to individuals with the highest degree appear first in the list. The uniformly random order within values of z ensures that the order varies among equally important links and across repeated simulations. See Figure 2.4 for an illustration of the sorting method.

This approach works well with the Add Health data and my simulated networks because the distribution of neighborhood sizes is fairly compact, but the sorting method may lead to lists with little variation when only a handful of individuals are likely to be observed for any given neighborhood size. In that case, one could introduce a white noise component to the ordering. Since the ordering would then be a combination of the above order and links missing at random as in Chapter 1, I expect that in general the bias will tend to fall in between these two extremes.

Figure 2.4
Sorting ℓ to be positively correlated with neighbor's degree



In step (1) each node pair (i, j) is assigned a value z equal to g_j. In step (2) ℓ is sorted uniformly at random. In step (3) ℓ is sorted by z maintaining the uniformly random order within equal values of z. Although all elements are sorted, I only display arrows linking the first five elements in steps (2) and (3) for readability.

2.6.1 Application to Add Health

To estimate the effects of missing links when individuals are more likely to report links to well connected individuals, I conduct 200 Monte Carlo simulations on the Add Health sample. I estimate the mean and standard deviation of the local-aggregate and local-average effects across simulations with varying fractions of missing links. My results are reported in Table 2.7.

When reporting is correlated with neighbors degree centrality and a significant fraction of links are missing the local-aggregate effect can be inflated by more than 80 percent. The bias decreases as more links are observed but persists. When half of the network is observed the bias is still over 50 percent. Unbiased estimates of the local-aggregate effect are only possible when at least 85 percent of all links are observed.

The local-average effect's bias is even more erratic. The estimated effect on recreational activities bares the wrong sign when less than 25 percent of links are observed but also overestimates the baseline estimate by more than 28 percent when 80 percent of links are observed. The estimated effect on physical activities is also initially of the wrong sign before climbing to more than double the baseline value. Only when all links are observed is the local-average effect unbiased. The results stand in stark contrast to links missing at random that almost everywhere generate an attenuating effect and may be negligible when most of the network is observed.

2.6.2 Robustness check

As in Section 5.3, I also estimate the effects of links missing not at random on a set of Erdős-Rényi random graphs. I use the same parameters here as in Section 5.3 above. Table 2.8 reports the mean and standard deviation of the estimated local-aggregate effect. Like the Add Health data, non-random missing links inflate estimates of the local-aggregate effect. However, the bias is far less persistent. Contrary to what one might expect the bias is more persistent with larger values of λ_1. This can be seen by comparing the mean estimates of λ_1 when 50 percent of links are observed. The estimated effect is unbiased when λ_1 equals 0.01, inflated by 10 percent when λ_1 equals 0.05 and inflated by 14 percent when λ_1 equals 0.1.

The effect on the local-average estimates, reported in Table 2.9, appear more in line with what one might expect based on other forms of measurement error. The local-average effect is downward biased and increasing in the fraction of observed links. Unlike the Add Health data, the bias is never inflationary. Unfortunately, the local-average estimates are only unbiased when all links are observed.

2.7 Spurious links

Thus far, I have focused on the absence of links either because of survey induced top-coding or because of mis-reporting that is correlated with neighborhood size. I now turn to the problem of spurious links, that is links that are observed but do not reflect a genuine interaction. For example, a student might report being friends with a popular student or a crush when no meaningful relationship actually exists.

In order to isolate the effect of spurious links I restrict my attention to scenarios where spurious links are the only form of misspecification. To that end, I begin with the observed network and add spurious links to it. I assume that links are spurious at random, i.e. the likelihood of observing any particular spurious link is as likely as any other.

The set of spurious links is given by the difference between the set of all possible links and the set of links in ℓ. The set of all possible links contains $n(n-2)$ elements and can be constructed by taking the cartesian product of the list of all n individuals with itself and removing self-links, i.e. links of the form (i, i). I call the set of all possible links $\mathbb{L}$.

Modeling spurious links is then a matter of choosing a set of spurious links from $\hat{\ell} = \mathbb{L} \backslash \ell$ and estimating the effect of adding $\hat{\ell}$ links to the existing network. The set of spurious links that is added to the given network is chosen by sorting the set of all potentially spurious links, $\hat{\ell}$, uniformly at random and taking the first $|\ell|/2$, i.e. the total number of spurious links to be added is half the size of the total number of links in ℓ. Estimation then proceeds as follows. First, construct the observed network from the list ℓ and derive each of the matrices in equation (2.1). Next add links one at a time from the set of spurious links, $\hat{\ell}$, to the existing network. Estimates of equation (2.1) based on the networks constructed with spurious links provide an empirical measure of the bias caused by adding those links. Repeating the process with many independently selected sets of spurious links allows one to estimate the empirical distribution of the bias.

2.7.1 Application to Add Health

Table 2.10 reports the mean and standard deviation of estimates of equation (2.1) as spurious links are added to the Add Health sample. The mean local-aggregate effect of neighbors recreational and physical activity is decreasing in the number of spurious links. The bias is modest amounting to a less than 10 percent reduction in the estimated effect on recreational activities after the total number of links has increased by 25 percent and less than 19 percent for physical activities. The relationship appears to be approximately linear in the vicinity of the baseline estimates with the bias roughly doubling as the number of spurious links doubles.

The local-average bias is almost everywhere increasing in the number of spurious links. The bias in the more precisely estimated effect of recreational activities accounts for a less than 8 percent increase when the total number of links has increased by 25 percent and less than 15 percent after the total number of links has increased by 50 percent. The bias in physical activities is larger, an approximately 43 percent increase, after the total number of links has increased by 50 percent.

Robustness checks

The results from Erdős-Rényi random graphs contrast with the Add Health findings. The estimated bias for the both the local-aggregate and local-average effects trends almost everywhere toward zero in the random graphs. The only exception is when λ_1 equals 0.01, where estimates are inflated by less than 10 percent when links are added until the total number of links has increased by 50 percent and the bias disappears. The bias for all other values of λ_1 and λ_2 is a roughly linear downward trend in both the local-aggregate and local-average effects that amounts to a less than 5 percent reduction in the estimated effect for every 5 percent of additional links.

2.8 Discussion

This paper extends the methods in Chapter 1 to study the effects of three previously under explored sources of measurement error: top-coding neighborhood size, mis-reporting that is correlated with the network topology and spurious links. In addition, I provide empirical evidence on the effects of each through Monte Carlo simulations using the Add Health survey and Erdős-Rényi random graphs. Overall, I find that mis-reporting that is correlated with the network topology is the most concerning form of measurement error. This type of mis-reporting can lead to overestimates of both local-aggregate and local-average effects. Furthermore, even when almost the entire network is observed the local-average effect is still likely to be biased. To my knowledge this is the first paper to estimate the effects of non-random mis-reporting.

My findings also complement those of Griffith (2019) and De Giorgi et al. (2010). Whereas Griffith focuses on top-coding in a linear-in-means model, I estimate the effect on a generalized model that permits estimates of local-aggregate and local-average effects in addition to exogenous effects. I find that unlike the exogenous effect the local-aggregate and local-average effects may be biased either up or down. The effect is typically modest when the limit approaches the maximum neighborhood size.

De Giorgi et al. (2010) estimate the effects of mis-reporting in a local-average model where links may be either missing or spurious. To avoid confounding with links missing at random I isolate the effects of spurious links. I find that spurious links exert a downward bias on estimates of the local-aggregate effect. The change in estimates of the local-average effect is less clear. In the Add Health data spurious links generate an upward bias and in the Erdős-Rényi random graphs spurious links produce a downward bias. If the bias caused by spurious links were in the opposite direction of the bias caused by missing links in De Giorgi et al. (2010) it could help explain why the authors reported very little bias in there study.

To my knowledge this is the first paper to report on the effects of top-coding in a local-aggregate or local-average model, to discuss the effects of mis-reporting that is correlated with the network topology or to attempt to isolate the effects of spurious links. While my findings do provide additional evidence for the relative robustness of local-aggregate models to various forms of measurement error, they also highlight that these models are not immune to measurement error. In addition, differences between the bias in the Add Health data and the Erdős-Rényi random graphs suggest that factors beyond the type of misspecification may contribute to biased estimators. Therefore, I recommend that applied researchers apply these techniques to test the sensitivity of their models on their own data.

Table 2.1
Descriptive statistics

Variable	Mean	Std. Dev.
Recreational activities	2.309	1.458
Physical activity	2.373	1.191
Age	14.934	1.743
Female	0.510	0.500
White only	0.698	0.459
Mom and Dad present at home	0.749	0.434
Mom has a college degree	0.220	0.414
Dad has a college degree	0.199	0.399
Mom works	0.737	0.441
Dad works	0.724	0.447
Number of observations 3,061		

Data was obtained from a sample of ten schools with the lowest reported error rates among friendship nominations drawn from the Add Health Wave 1 In School questionnaire. Recreational activity is measured by an index of participation in extracurricular activities such as clubs and sports with an upper limit of four as in Bramoullé et al. (2009). Physical activity is measured by the question, "How many times in a normal week do you work, play or exercise hard enough to make you sweat and breath heavily?", coded as never (0), one or two times (1), three to five times (2), six or seven times (3) and more than seven times (4) as in Liu et al. (2014). The unit of observation is an individual.

Table 2.2
Estimation of peer effects in recreational activities

	Variable	Coefficient	SE
Local-aggregate effect (λ_1)	Recreational activities	0.025***	0.004
Local-average effect (λ_2)	Recreational activities	0.468***	0.092
Own-characteristics (β)	Age	0.010	0.017
	Female	0.280***	0.051
	Race is white only	-0.170**	0.078
	Mom and dad present	-0.024	0.093
	Mom has college degree	0.214***	0.063
	Dad has college degree	0.250***	0.066
	Mom works	0.103*	0.057
	Dad works	0.124	0.089
Average friends' characteristics (γ)	Age	-0.066***	0.012
	Female	-0.117	0.106
	Race is white only	-0.171	0.121
	Mom and dad present	0.006	0.194
	Mom has college degree	0.210	0.140
	Dad has college degree	0.393***	0.152
	Mom works	0.028	0.123
	Dad works	0.002	0.182

Effect of peer recreational activity on own-activity. Estimates based on the bias corrected 2SLS method proposed by Liu et al. (2014). The unit of observation is an individual. Data was obtained from a sample of ten schools with the lowest reported error rates among friendship nominations drawn from the Add Health Wave 1 In School questionnaire. Recreational activity is measured by an index of participation in extracurricular activities such as clubs and sports with an upper limit of four as in Bramoullé et al. (2009). Standard errors are heteroscedasticity robust. Statistical significance: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 2.3
Estimation of peer effects in physical activities

	Variable	Coefficient	SE
Local-aggregate effect (λ_1)	Physical activities	0.011***	0.003
Local-average effect (λ_2)	Physical activities	0.169	0.109
Own-characteristics (β)	Age	-0.049**	0.021
	Female	-0.709***	0.043
	White only	0.058	0.065
	Mom and dad present	0.106	0.078
	Mom college degree	0.133**	0.052
	Dad college degree	-0.002	0.055
	Mom works	-0.025	0.047
	Dad works	-0.006	0.075
Average friends' characteristics (γ)	Age	-0.037**	0.015
	Female	0.144	0.103
	White only	-0.144	0.112
	Mom and dad present	-0.075	0.163
	Mom college degree	0.125	0.112
	Dad college degree	0.060	0.116
	Mom works	-0.052	0.099
	Dad works	0.160	0.153

Effect of peer physical activity on own-activity. Estimates based on the bias corrected 2SLS method proposed by Liu et al. (2014). The unit of observation is an individual. Data was obtained from a sample of ten schools with the lowest reported error rates among friendship nominations drawn from the Add Health Wave 1 In School questionnaire. Physical activity is measured by the question, "How many times in a normal week do you work, play or exercise hard enough to make you sweat and breath heavily?", coded as never (0), one or two times (1), three to five times (2), six or seven times (3) and more than seven times (4) as in Liu et al. (2014). Standard errors are heteroscedasticity robust. Statistical significance: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 2.4
Peer effects with top-coding in the Add Health data

Recreational activities					Physical activities				
Limit	λ_1	SD	λ_2	SD	Limit	λ_1	SD	λ_2	SD
1	0.124	0.010	0.128	0.095	1	0.022	0.004	0.115	0.061
2	0.047	0.003	0.403	0.065	2	0.017	0.001	0.139	0.045
3	0.034	0.001	0.445	0.037	3	0.013	0.000	0.178	0.034
4	0.028	0.001	0.455	0.020	4	0.011	0.000	0.175	0.020
5	0.025	-	0.468	-	5	0.011	-	0.169	-

Monte Carlo simulations of the effect of peer recreational and physical activity on own-activity when neighborhoods are top-coded. Data was obtained from a sample of ten schools with the lowest reported error rates among friendship nominations drawn from the Add Health Wave 1 In School questionnaire. Recreational activity is measured by an index of participation in extracurricular activities such as clubs and sports with an upper limit of four as in Bramoullé et al. (2009). Physical activity is measured by the question, "How many times in a normal week do you work, play or exercise hard enough to make you sweat and breath heavily?", coded as never (0), one or two times (1), three to five times (2), six or seven times (3) and more than seven times (4) as in Liu et al. (2014). The unit of observation is an individual, $n = 3,061$ with $\ell = 16,777$ links. $B = 200$ repetitions were performed for each outcome.

Table 2.5
Local-aggregate effects with top-coding in Erdős-Rényi random graphs

	$[\lambda_1 = 0.01]$		$[\lambda_1 = 0.05]$		$[\lambda_1 = 0.10]$	
Limit	Mean	SD	Mean	SD	Mean	SD
1	0.372	0.046	0.549	0.063	0.993	0.122
2	-0.025	0.007	0.128	0.01	0.311	0.018
3	0.008	0.003	0.102	0.003	0.216	0.006
4	0.013	0.001	0.080	0.001	0.161	0.003
5	0.013	0.001	0.068	0.001	0.135	0.001
6	0.012	0.001	0.060	0.001	0.119	0.001
7	0.012	0.000	0.055	0.000	0.109	0.000
8	0.011	0.000	0.052	0.000	0.104	0.000
9	0.011	0.000	0.051	0.000	0.101	0.000
10	0.011	-	0.051	-	0.100	-

Monte Carlo simulations of the effect of peer outcomes on own-outcomes when neighborhoods are top-coded. Each network contains five Erdős-Rényi random graphs with $n = 1,000$ individuals and $\ell = 4,000$ links distributed across five equally sized clusters. Mean and standard deviations based on $B = 500$ repetitions with $\lambda_2 = 0.5$, $\beta = 0.25$, $\gamma = 0.15$, $\delta = 1$ and $\sigma^2 = 0.009$ held constant while varying λ_1.

Table 2.6
Local-average effects with top-coding in Erdős-Rényi random graphs

% of links	$[\lambda_2 = 0.1]$ Mean	SD	$[\lambda_2 = 0.3]$ Mean	SD	$[\lambda_2 = 0.5]$ Mean	SD
1	-	-	-	-	-	-
2	0.199	0.043	0.311	0.046	0.436	0.052
3	0.095	0.031	0.251	0.032	0.416	0.035
4	0.069	0.019	0.251	0.019	0.439	0.021
5	0.062	0.013	0.255	0.013	0.449	0.013
6	0.069	0.007	0.266	0.007	0.463	0.008
7	0.082	0.004	0.282	0.004	0.482	0.004
8	0.090	0.002	0.291	0.002	0.492	0.002
9	0.092	0.001	0.293	0.001	0.494	0.001
10	0.092	-	0.294	-	0.495	-

Monte Carlo simulations of the effect of peer outcomes on own-outcomes when neighborhoods are top-coded. Each network contains five Erdős-Rényi random graphs with $n = 1,000$ individuals and $\ell = 4,000$ links distributed across five equally sized clusters. Mean and standard deviations based on $B = 500$ repetitions with $\lambda_1 = 0.05$, $\beta = 0.25$, $\gamma = 0.15$, $\delta = 1$ and $\sigma^2 = 0.009$ held constant while varying λ_2.

Table 2.7
Peer effects with non-random mis-reporting in the Add Health data

	Recreational activities					Physical activities			
% of	$[\lambda_1]$		$[\lambda_2]$		% of	$[\lambda_1]$		$[\lambda_2]$	
links	Mean	SD	Mean	SD	links	Mean	SD	Mean	SD
5	0.045	0.019	-0.261	0.121	5	0.000	0.023	-0.185	0.108
10	0.043	0.003	-0.083	0.043	10	0.007	0.003	0.157	0.044
15	0.045	0.004	-0.175	0.096	15	0.016	0.004	-0.037	0.113
20	0.044	0.003	-0.070	0.094	20	0.019	0.003	0.000	0.121
25	0.043	0.000	-0.047	0.012	25	0.020	0.000	0.029	0.013
30	0.042	0.002	0.109	0.081	30	0.018	0.002	0.053	0.082
35	0.040	0.002	0.238	0.071	35	0.017	0.002	-0.002	0.077
40	0.039	0.000	0.302	0.013	40	0.016	0.000	-0.103	0.020
45	0.038	0.002	0.428	0.059	45	0.017	0.002	0.030	0.063
50	0.037	0.001	0.488	0.044	50	0.017	0.001	0.106	0.059
55	0.037	0.001	0.475	0.057	55	0.017	0.001	0.176	0.065
60	0.035	0.001	0.495	0.052	60	0.016	0.001	0.283	0.073
65	0.033	0.001	0.532	0.026	65	0.016	0.001	0.329	0.046
70	0.031	0.001	0.550	0.044	70	0.015	0.001	0.346	0.058
75	0.029	0.000	0.598	0.016	75	0.015	0.000	0.363	0.024
80	0.027	0.001	0.602	0.032	80	0.014	0.001	0.451	0.066
85	0.025	0.001	0.584	0.032	85	0.012	0.001	0.442	0.047
90	0.024	0.001	0.551	0.019	90	0.011	0.000	0.373	0.029
95	0.023	0.001	0.476	0.024	95	0.011	0.000	0.196	0.037
100	0.025	-	0.468	-	100	0.011	-	0.169	-

Monte Carlo simulations of the effect of peer recreational and physical activity on own-activity with non-random mis-reporting. Observing a link is positively correlated with the degree of the neighbor the link points to. Data was obtained from a sample of ten schools with the lowest reported error rates among friendship nominations drawn from the Add Health Wave 1 In School questionnaire. Recreational activity is measured by an index of participation in extracurricular activities such as clubs and sports with an upper limit of four as in Bramoullé et al. (2009). Physical activity is measured by the question, "How many times in a normal week do you work, play or exercise hard enough to make you sweat and breath heavily?", coded as never (0), one or two times (1), three to five times (2), six or seven times (3) and more than seven times (4) as in Liu et al. (2014). The unit of observation is an individual, $n = 3,061$ with $\ell = 16,777$ links. $B = 200$ repetitions were performed for each outcome.

Table 2.8
Local-aggregate effects with non-random mis-reporting in Erdős-Rényi random graphs

% of links	$[\lambda_1 = 0.01]$ Mean	SD	$[\lambda_1 = 0.05]$ Mean	SD	$[\lambda_1 = 0.10]$ Mean	SD
5	0.125	0.026	0.121	0.027	0.120	0.028
10	0.074	0.016	0.094	0.020	0.107	0.030
15	0.048	0.010	0.075	0.010	0.108	0.014
20	0.029	0.005	0.064	0.005	0.112	0.006
25	0.020	0.004	0.063	0.004	0.121	0.006
30	0.015	0.003	0.060	0.003	0.119	0.004
35	0.015	0.001	0.060	0.001	0.116	0.001
40	0.012	0.002	0.057	0.002	0.115	0.003
45	0.010	0.002	0.056	0.002	0.114	0.003
50	0.010	0.001	0.055	0.001	0.114	0.002
55	0.010	0.001	0.055	0.001	0.113	0.001
60	0.009	0.001	0.054	0.001	0.112	0.002
65	0.009	0.001	0.053	0.001	0.110	0.001
70	0.010	0.001	0.053	0.001	0.109	0.001
75	0.010	0.001	0.052	0.001	0.107	0.000
80	0.009	0.001	0.051	0.001	0.105	0.001
85	0.010	0.001	0.052	0.001	0.104	0.001
90	0.011	0.000	0.052	0.000	0.104	0.000
95	0.011	0.001	0.051	0.001	0.102	0.000
100	0.011	-	0.051	-	0.100	-

Monte Carlo simulations of the effect of peer outcomes on own-outcomes when observing a link is positively correlated with the degree of the neighbor the link points to.. Each network contains five Erdős-Rényi random graphs with $n = 1,000$ individuals and $\ell = 4,000$ links distributed across five equally sized clusters. Mean and standard deviations based on $B = 500$ repetitions with $\lambda_2 = 0.5$, $\beta = 0.25$, $\gamma = 0.15$, $\delta = 1$ and $\sigma^2 = 0.009$ held constant while varying λ_1.

Table 2.9
Local-average effects with non-random mis-reporting in Erdős-Rényi random graphs

% of	[$\lambda_2 = 0.1$]		[$\lambda_2 = 0.3$]		[$\lambda_2 = 0.5$]	
links	Mean	SD	Mean	SD	Mean	SD
5	-0.3	0.043	-0.292	0.044	-0.274	0.041
10	-0.322	0.045	-0.297	0.045	-0.264	0.044
15	-0.259	0.046	-0.238	0.043	-0.212	0.037
20	-0.177	0.04	-0.165	0.036	-0.166	0.028
25	-0.139	0.041	-0.123	0.04	-0.136	0.036
30	-0.081	0.045	-0.046	0.046	-0.052	0.048
35	-0.014	0.012	0.045	0.014	0.076	0.018
40	-0.042	0.038	0.016	0.041	0.048	0.045
45	-0.027	0.041	0.04	0.043	0.079	0.046
50	-0.003	0.037	0.074	0.037	0.12	0.04
55	0.022	0.019	0.104	0.02	0.156	0.023
60	0.027	0.035	0.11	0.037	0.157	0.042
65	0.039	0.034	0.134	0.036	0.19	0.04
70	0.055	0.026	0.165	0.027	0.241	0.032
75	0.058	0.017	0.181	0.018	0.274	0.022
80	0.057	0.025	0.193	0.026	0.304	0.031
85	0.07	0.024	0.227	0.026	0.37	0.031
90	0.083	0.014	0.262	0.015	0.436	0.018
95	0.069	0.014	0.261	0.016	0.451	0.021
100	0.092	-	0.294	-	0.495	-

Monte Carlo simulations of the effect of peer outcomes on own-outcomes when observing a link is positively correlated with the degree of the neighbor the link points to.. Each network contains five Erdős-Rényi random graphs with $n = 1,000$ individuals and $\ell = 4,000$ links distributed across five equally sized clusters. Mean and standard deviations based on $B = 500$ repetitions with $\lambda_2 = 0.5$, $\beta = 0.25$, $\gamma = 0.15$, $\delta = 1$ and $\sigma^2 = 0.009$ held constant while varying λ_1.

Table 2.10
Peer effects with spurious links in the Add Health data

	Recreational activities					Physical activities			
% of	$[\lambda_1]$		$[\lambda_2]$		% of	$[\lambda_1]$		$[\lambda_2]$	
links	Mean	SD	Mean	SD	links	Mean	SD	Mean	SD
0	0.025	-	0.468	-	0	0.011	-	0.169	-
5	0.025	0.002	0.463	0.055	5	0.010	0.001	0.194	0.040
10	0.025	0.002	0.477	0.064	10	0.010	0.001	0.210	0.051
15	0.024	0.002	0.486	0.073	15	0.009	0.001	0.216	0.057
20	0.024	0.002	0.495	0.079	20	0.009	0.001	0.224	0.057
25	0.023	0.002	0.502	0.083	25	0.009	0.001	0.227	0.064
30	0.022	0.002	0.504	0.088	30	0.008	0.001	0.227	0.066
35	0.021	0.002	0.514	0.088	35	0.008	0.001	0.230	0.067
40	0.021	0.002	0.520	0.093	40	0.008	0.001	0.234	0.069
45	0.020	0.002	0.528	0.097	45	0.007	0.001	0.237	0.072
50	0.019	0.002	0.534	0.096	50	0.007	0.001	0.239	0.075

Monte Carlo simulations of the effect of peer recreational and physical activity on own-activity when links are spurious at random. Data was obtained from a sample of ten schools with the lowest reported error rates among friendship nominations drawn from the Add Health Wave 1 In School questionnaire. Recreational activity is measured by an index of participation in extracurricular activities such as clubs and sports with an upper limit of four as in Bramoullé et al. (2009). Physical activity is measured by the question, “How many times in a normal week do you work, play or exercise hard enough to make you sweat and breath heavily?”, coded as never (0), one or two times (1), three to five times (2), six or seven times (3) and more than seven times (4) as in Liu et al. (2014). The unit of observation is an individual, $n = 3,061$ with $\ell = 16,777$ links. $B = 200$ repetitions were performed for each outcome.

Table 2.11
Local-aggregate effects with spurious links in Erdős-Rényi random graphs

% of spurious links	$[\lambda_1 = 0.01]$ Mean	SD	$[\lambda_1 = 0.05]$ Mean	SD	$[\lambda_1 = 0.10]$ Mean	SD
0	0.011	-	0.051	-	0.1	-
5	0.012	0.001	0.049	0.001	0.097	0.001
10	0.012	0.001	0.048	0.001	0.094	0.002
15	0.012	0.002	0.047	0.001	0.090	0.002
20	0.012	0.002	0.046	0.002	0.087	0.002
25	0.012	0.002	0.044	0.002	0.084	0.002
30	0.012	0.002	0.043	0.002	0.082	0.002
35	0.012	0.002	0.042	0.002	0.079	0.002
40	0.012	0.002	0.041	0.001	0.077	0.002
45	0.012	0.002	0.039	0.001	0.074	0.002
50	0.011	0.001	0.038	0.001	0.072	0.002

Monte Carlo simulations of peer effects with spurious links. Each network contains five Erdős-Rényi random graphs with $n = 1,000$ individuals and $\ell = 4,000$ links distributed across five equally sized clusters. Mean and standard deviation based on $B = 500$ repetitions with $\lambda_2 = 0.5$, $\beta = 0.25$, $\gamma = 0.15$, $\delta = 1$ and $\sigma^2 = 0.009$ held constant while varying λ_1.

Table 2.12
Local-average effects with spurious links in Erdős-Rényi random graphs

% of spurious links	$[\lambda_2 = 0.1]$ Mean	SD	$[\lambda_2 = 0.3]$ Mean	SD	$[\lambda_2 = 0.5]$ Mean	SD
0	0.092	-	0.294	-	0.495	-
5	0.090	0.018	0.289	0.021	0.488	0.027
10	0.086	0.026	0.282	0.030	0.477	0.042
15	0.082	0.034	0.276	0.040	0.469	0.055
20	0.080	0.039	0.273	0.046	0.463	0.064
25	0.078	0.044	0.269	0.051	0.457	0.070
30	0.076	0.049	0.265	0.057	0.449	0.080
35	0.075	0.055	0.263	0.065	0.446	0.093
40	0.073	0.062	0.260	0.072	0.442	0.101
45	0.073	0.065	0.258	0.077	0.438	0.110
50	0.072	0.071	0.257	0.083	0.434	0.119

Monte Carlo simulations of peer effects with spurious links. Each network contains five Erdős-Rényi random graphs with $n = 1.000$ individuals and $\ell = 4.000$ links distributed across five equally sized clusters. Mean and standard deviation based on $B = 500$ repetitions with $\lambda_1 = 0.05$, $\beta = 0.25$, $\gamma = 0.15$, $\delta = 1$ and $\sigma^2 = 0.009$ held constant while varying λ_2.

CHAPTER 3

NURSE OUTREACH AND FREQUENT EMERGENCY DEPARTMENT USERS: A SYNTHETIC CONTROL ANALYSIS

3.1 Introduction

The United States spends a staggering 17.5 percent of GDP on health care with an expected average annual growth rate of 5.6 percent over the next decade (Keehan et al., 2017). The rapid growth of health care spending has strained public programs, such as Medicare and Medicaid, with Medicare now expected to deplete the Hospital Insurance trust fund by 2026 (The Board of Trustees and Funds, 2021). Moreover, as much as 30 percent of Medicare spending is wasted on inefficient expenditures (McGinnis et al., 2013). The National Academy of Medicine estimates that nearly half of superfluous spending is due to unnecessary and inefficient services while fraud, missed prevention and higher prices drive the remaining waste (McGinnis et al., 2013).

Efforts like cost-sharing that aim to curb wasteful spending by better aligning prices with value have resulted in unclear and often context specific outcomes. Perhaps, because relying solely on prices to augment consumer behavior largely ignores providers, misses much of the consumers optimization problem. Health care consumers face complex and uncertain decisions over a set of often uncomfortable and risky alternatives, while providers can be incentivized to offer more care.

The decision to go to a hospital emergency department in lieu of comparable care in a less costly setting presents a prime example of the disconnect between price and health care demand. Emergency departments are one of the most expensive places to receive care, costing hospitals a total of $76.3 billion to provide care in 2017 (Moore and Liang, 2020) and charging patients far more than alternative settings for similar care (Insinga et al., 2011). Patients know that the emergency department is an expensive source of care (Newton et al., 2008), yet emergency department utilization has grown in the last decade, and overcrowding is a persistent concern (Newton et al., 2008).

Policy makers, payers and providers have been keen to develop systems that alleviate costly inefficiencies in the delivery of emergency care. In response, programs that have shown promise in other settings, have adapted to assist emergency department patients by providing decision support, referrals, entitlement assistance and care coordination (Kumar and Klein, 2013).[1] Although a number of studies suggest that such programs can reduce emergency department utilization, the simultaneous adoption and heterogenous application of interventions has

[1]Referrals may be to primary care, substance abuse, pain, psychiatric or social services.

hindered researchers ability to pinpoint the specific aspects of these programs that are most cost effective (Van den Heede and Van de Voorde, 2016).

To gain insight into the efficacy of a specific intervention, I analyze the effects of a nurse outreach program administered by a major insurance carrier throughout 2013. The primary goal of the program is to promote primary care engagement among currently insured Medicare Supplement beneficiaries with a history of frequent emergency department use. Ideally, this would lead to a shift away from more costly and intensive care in an emergency room or hospital toward less costly care in a primary care physicians office.

Program participants are contacted over the phone by a registered nurse. The nurse reviews the individual's medical history and verifies that they have a primary care physician. If the individual does not, the nurse helps them find one. The nurse can then facilitate scheduling appointments and obtaining referrals as needed. Finally, the nurse discusses treatment options before referring participants to a toll free telephone line staffed by registered nurses twenty-four hours a day who can evaluate symptoms, answer questions about medications and direct callers to appropriate health care.

I employ synthetic control methods to estimate the impact of nurse outreach on health care utilization. Six months after enrollment, I find that treated individuals' primary care utilization increased by roughly 15 percent on average compared to the synthetic control group. This difference is statistically significant and robust to placebo tests. However, I do not find clear evidence for an effect of treatment on emergency department utilization. The estimates appear to be sensitive to the model specification and are most often not statistically different from zero. My findings indicate that nurse outreach did increase primary care engagement, but did not produce a detectable effect on emergency department utilization, undermining the case for any potential cost savings.

The rest of the paper is organized as follows. Section 2 discusses previous findings on interventions for frequent emergency department utilization. Section 3 describes the synthetic control methods used for analysis. Section 4 details the intervention and data set. Section 5 presents my findings, and Section 6 concludes.

3.2 Background

There is a large body of medical literature concerned with interventions for frequent emergency department users (Althaus et al., 2011; Burns et al., 2007; Freeman, 2006; Kumar and Klein, 2013; Norris et al., 2002; Wells et al., 2008; Ziguras and Stuart, 2000). Although the literature does not provide a clear definition of "frequent use," the literature does find that recurrent users consume a disproportionate share of health care resources, tend to have multiple co-morbid conditions and suffer from addiction and psychosocial issues at a higher rate than the general populace (Kumar and Klein, 2013; Soril et al., 2015). The research also suggests that receiving the right care at the right time may be a primary issue for frequent users.

One of the most commonly cited interventions for frequent emergency department users is case management. Case management is broadly defined as a comprehensive, interdisciplinary approach to improve outcomes by designating a single point of contact who helps assess, plan, personalize and guide an individual through a customized care process (Soril et al., 2015). Systematic reviews of the case management literature report a great deal of heterogeneity across studies. The definition of "frequent" ranges from as few as three visits per year to five or more visits per month (Kumar and Klein, 2013). The interventions also vary in intensity with some interventions relying on a single social worker to provide periodic outreach while others employ multi-disciplinary teams that can reach out as often as needed. The interventions are often multifaceted and can include assistance with primary care engagement, crisis intervention, care coordination, entitlements and referrals (Althaus et al., 2011; Soril et al., 2015). Although the amount of variation across studies makes direct comparisons difficult, the intensity of treatment and the appropriateness of the target population (i.e. risk of frequent utilization combined with the complexity of care needs) were associated with the greatest likelihood of reducing emergency room utilization (Soril et al., 2015; Hudon et al., 2019). The evidence on cost savings, however, remains unclear. Most studies do not report cost information, and the handful that do provide mixed results.

3.2.1 Cost-sharing

Much of the economics literature has focused on the effect of increased out of pocket costs on emergency department usage[2] (Choudhry et al., 2012; Hirth et al., 2016; Hsu et al., 2006; Nair et al., 2010; Siddiqui et al., 2015). Choudhry et al. (2012) and Nair et al. (2010) consider the impact of reducing co-payments for preventive medications prescribed for the treatment of cardiovascular disease and diabetes, respectively. Using pre-post study designs at large employers both studies find statistically significant reductions in emergency department utilization. However, both studies find that these effects are modest and neither study establishes a clear link between these interventions and appreciable cost-savings.

Siddiqui et al. (2015) studies the effect of co-payments for non-urgent emergency department utilization among Medicaid beneficiaries during the 2008 to 2010 recession. Eight states adopted co-payments under the Deficit Reduction Act of 2005. Using a differences-in-differences design the authors compare outcomes before and after the co-payments were implemented between the eight states enacting co-payments and ten other states with no co-payments. The authors found

[2] These studies use basic demand theory and the principles of value based insurance design to steer patients away from "low" value" services by charging higher out of pocket costs or towards "high value" services by charging less.

no statistically significant changes in emergency department utilization, primary care visits or inpatient days following the Deficit Reduction Act.

However, Hirth et al. (2016) and Hsu et al. (2006) do find statistically significant reductions in emergency department usage after increasing cost-sharing. Hirth et al. (2016) leverages the Health Enhancement Program implemented for Connecticut state employees in 2011 using a differences-in-differences design. The voluntary program followed value-based insurance design principles: lowering out of pocket costs for high value preventive services and imposing a surcharge for individuals not participating or not complying with the initiative. Hirth et al. (2016) find that primary care visits increased 13.5 percent in year one and 4.8 percent in year two and ER visits without a resulting inpatient stay decreased by 10 visits per 1,000 enrollees in year one and by 25 visits per 1,000 enrollees in year two.

Hsu et al. (2006) analyzes a quasi-experimental longitudinal study conducted by Kaiser Permanente-Northern California (KPNC). In January 2000, roughly half of the KPNC population experienced an increase in ER co-payments. Among the commercially insured, ER visits decreased 12 percent for the \$20-35 co-payment and 23 percent for the \$50-100 co-payment compared with no co-payment. Among Medicare subjects ER visits decreased by 4 percent for the \$20-50 co-payment compared to no co-payment without increasing unfavorable events.

However, neither Hirth et al. (2016) nor Hsu et al. (2006) find statistically significant savings from their respective cost-sharing programs. The extent to which cost-sharing can produce cost-savings remains unclear and may be context specific.

Unlike the treated groups in Hsu et al. (2006), Siddiqui et al. (2015) and Hirth et al. (2016), the target population in this paper has almost no financial incentive to seek the most cost effective care. Medicare supplement plans cover most co-payments and coinsurance not covered by Medicare. So that primary care must provide better quality and continuity of care than more expensive alternatives in order to be preferable.

3.2.2 Primary care engagement

One element of the health care delivery system that has garnered independent investigation is primary care engagement. Primary care engagement is often measured by the frequency with which patients interact with their primary care physician. Higher levels of primary care engagement were shown to reduce emergency department utilization, improve adherence with prevention and screening services and lower costs (Rosenblatt et al., 2000; Blewett et al., 2008; Sabety et al., 2021).

In order to be engaged patients first need to have a primary care physician. Navratil-Strawn et al. (2014) estimate the return on investment for a program that helps frequent emergency department users find a primary care physician and schedule appointments and referrals. The authors compare AARP branded Medicare Supplement plan holders with three or more emergency department visits in the last year who participated in the program with eligible plan holders who did not participate using a differences-in-difference with propensity score matching model. The authors find that participants were less likely to use the emergency department or be admitted to the hospital and were more likely to seek primary care than non-participants during the post-intervention period. However, the study did not find a statistically significant difference in participants and non-participants claim costs.

3.2.3 Contributions

Like Navratil-Strawn et al. (2014) I study the effects of a nurse outreach program designed to encourage primary care engagement among Medicare Supplement beneficiaries with a history of frequent emergency department utilization. A central challenge in both studies is identifying a suitable control group. I apply synthetic control methods to address this issue. Similar to propensity score matching synthetic controls assign non-negative weights to a target control group. However, rather than simply averaging a fixed number of controls, synthetic controls select a weighted average that most closely matches the treatment group (Abadie and LHour, 2021). To my knowledge, this is the first application of synthetic control methods to study an intervention for frequent emergency department users.

To facilitate my analysis, I obtained access to a unique sample of commercial claims data. The data allow me to match on individuals' pre-intervention emergency, primary and inpatient care utilization. Access to the data was granted as part of a data use agreement for the purpose of studying the effects of nurse outreach on health care utilization.

3.2.4 Program description

Eligibility for the nurse outreach program was determined based on emergency department utilization in the prior year. Due to the limited availability of nurses, from January to July 2013 individuals with three or more emergency department visits were eligible to participate and from August to December 2013 two or more emergency department visits were sufficient for eligibility.

Eligible individuals were contacted by phone. Individuals that chose to participate spoke with a registered nurse about their health status and treatment options. The nurse verified that the individual had a primary care physician. If the individual did not already have a primary care physician the nurse would help them find a physician. The nurse could then assist with scheduling appointments and obtaining referrals as needed. Finally, the nurse would refer participants to a toll free telephone line staffed by nurses who could evaluate symptoms, answer questions about medications and home care and direct callers to appropriate health care.

3.3 Data

In order to analyze the effects of nurse outreach I obtained a unique sample of commercial claims data through a data use agreement with a major U.S. insurer. The sample includes observations on 967 individuals that participated in the program in 2013 and 5,000 individuals who were never eligible. After excluding individuals with less than 18 months of pre-treatment data the final sample includes 895 treated individuals and 4,550 potential controls.

The sample includes individuals from all 50 states and the District of Columbia. All individuals were age 65 or older and enrolled in a Standardized Medicare Supplement plan.[3] Detailed claim-level data was used to construct measures of health care utilization and co-morbidity. The claim data was then aggregated to the individual-month level and anonymized in accordance with the data use agreement.[4]

I estimate the impact of the initial nurse outreach on two time-varying health outcomes: emergency department utilization and primary care visits.[5,6] The syn-

[3]The Centers for Medicare and Medicaid Services (CMS) standardized Medicare Supplement plan offerings into to ten plans, labeled A through N, with the passing of the Omnibus Budget Reconciliation Act of 1990.

[4]The data use agreement limits the use of Protected Health Information (PHI) as defined by the Health Insurance Portability and Accountability Act (HIPPA) as well as the company name. Individual identifiers are anonymized with a standard algorithm. Date identifiers are anonymized by calculating all dates relative to the program eligibility date and then masking the eligibility date.

[5]I define an emergency department visit as a claim with a unique date of service and either a revenue code of 450, 451, 452, 453, 454, 455, 457, 458, 459 or 981 or a procedure code of 99281 99282, 99283, 99284, 99285, 99286, 99287, 99288, 99291 or 99292. Revenue codes provide the preferred method for identifying emergency department use; however, revenue codes are not available on all claims in the data set. When a revenue code is not available I rely on the procedure code instead. This may overstate the number of emergency department visits in some cases.

[6]I define a primary care visit as a claim with a unique date of service and a procedure code of 99201, 99202, 99203, 99204, 99205, 99206, 99207, 99208, 99209, 99210, 99211, 99212, 99213,

thetic control group is constructed to match the treated group's aggregate pre-treatment emergency room, primary care and inpatient hospital utilization.[7] As shown in Robbins et al. (2017), the potential bias in equation (3.4) is expected to decrease as the number of outcomes and/or the number of pre-intervention time periods grows.

I also match on four time-invariant covariates: age, gender, rurality and co-morbidity. Age, gender and rurality are measured in the month prior to treatment. Rurality is a proxy for access. Rural communities may have fewer readily available health care providers leading to important implications for the way individuals engage with the health care system. I base my rurality indicator on metropolitan core-based statistical areas, as defined by the Census Bureau. I measure co-morbidity with the Charleson co-morbidity index. The Charleson index is a commonly used indicator of disease burden and predictor of one-year mortality (D'Hoore et al., 1996). The index is constructed from claim-level diagnostic data captured during the year prior to treatment.

3.4 Methods

Comparative case studies like those discussed above often rely on matching each treated unit with a single control as in propensity score matching or a small number of controls as in differences-in-differences. Synthetic control methods use a large donor pool of potential controls to compute a weighted sample that most closely matches the treatment group. A number of synthetic control methods have

99214, 99215, 99241, 99242, 99243, 99244, 99245, 99381, 99382, 99383, 99384, 99385, 99386, 99387, 99388, 99389, 99390, 99391, 99392, 99393, 99394, 99385, 99396, 99397 or 99432.

[7] I define an inpatient hospital admission as a claim with a unique date of service and a revenue code of 100, 101, 103, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146,147, 148, 149, 150, 151, 152, 153, 154, 155, 156,157, 158, 159, 160, 164, 167, 169, 190, 191, 192,193, 194, 199, 200, 201, 202, 203, 204, 206, 207, 208, 209, 210, 211, 212, 213, 214 or 219.

been proposed in the literature. I utilize the method proposed by Robbins et al. (2017) because it allows for disaggregated data with a large number of treated units and multiple outcome variables.[8] The setting for this method is outlined in the following section.

3.4.1 Setting

Suppose data is available on J total units. Index units so that $\{1, \ldots, J_0\}$ are untreated and $\{J_0 + 1, \ldots, J\}$ receive the intervention. Similarly, index time periods so that $\{1, \ldots, T_0\}$ correspond to the pre-intervention period and $\{T_0 + 1, \ldots, T\}$ correspond to the post-intervention period. The data includes I time-varying outcomes of interest and L time-invariant covariates. Denote the value of outcome i for unit j at time t by Y_{ijt} and the value of covariate ℓ for unit j by $X_{\ell j}$.

The objective is to estimate the effect of the intervention. Define $Y_{ijt}(0)$ to be the treatment-free potential outcome and $Y_{ijt}(1)$ to be the observed outcome with treatment in the post-intervention period. Prior to T_0 the treatment-free potential outcome, $Y_{ijt} = Y_{ijt}(0)$, is observable for all units. After T_0 only the outcome with intervention, $Y_{ijt} = Y_{ijt}(1)$, is observable for the treated units. The challenge then is to come up with an appropriate counterfactual outcome, $Y_{ijt}(0)$, for the treated units, $j \in (J_0 + 1, \ldots, J)$, in the post-intervention period, $t > T_0$.

Synthetic control methods seek to approximate the potential outcome $Y_{ijt}(0)$ for treated units in the post-intervention period with a weighted combination of untreated units. The method proposed by Robbins et al. (2017) consists of a set of weights w_j applied to the untreated units, $j \in \{1, \ldots, J_0\}$, that satisfy the following three sets of constraints. First, the sum of the weights equals the number

[8] The method proposed by Robbins et al. (2017) is implemented in the R package `microsynth`.

of treated cases, so that

$$\sum_{j=1}^{J_0} w_j = J - J_0. \tag{3.1}$$

Second, the synthetic control group's aggregate weighted covariates equal the treatment group's aggregate covariates. That is,

$$\sum_{j=1}^{J_0} w_j X_{\ell j} = \sum_{j=J_0+1}^{J} X_{\ell j} \quad \text{for } \ell \in (1, \ldots, L). \tag{3.2}$$

Third, each of the synthetic control group's aggregate weighted outcomes match each of the treatment group's aggregate outcomes in each pre-intervention time-period,

$$\sum_{j=1}^{J_0} w_j Y_{ijt} = \sum_{j=J_0+1}^{J} Y_{ijt} \quad \text{for all } i \in (1, \ldots, I) \text{ and } t \in (1, \ldots, T_0). \tag{3.3}$$

Taken together these constraints ensure that the synthetic control group is the same size as the treatment group and matches the treatment group's time-invariant covariates and time-varying outcomes.

When the treatment and donor pools are sufficiently large, finding weights $(w_1, \ldots, w_{J_0})$ that satisfy equations (3.1) - (3.3) exactly can be achieved with methods commonly used in the analysis of surveys. If an exact solution is infeasible, the constraints in equation (3.3) (and if necessary equation (3.2)) can be relaxed to permit a proximate solution.

Given weights $(w_1, \ldots, w_{J_0})$ satisfying equations (3.1) - (3.3), the post-intervention treatment-free potential outcome for the treated units can be approximated by $\hat{Y}_{it}(0) = \sum_{j=1}^{J_0} w_j Y_{ijt}$. And, the effect of the intervention may be estimated by averaging the difference between the treated units and the synthetic control, that is the treatment effect,

$$\hat{\alpha}_{it} = \frac{1}{J - J_0}\left(\sum_{j=J_0+1}^{J} Y_{ijt} - \sum_{j=1}^{J_0} w_j Y_{ijt}\right) \tag{3.4}$$

for outcome i in time period t when $t > T_0$. Averaging the treatment effect, $\hat{\alpha}_{it}$, over the post-intervention period yields an estimate of the average treatment effect on the treated,

$$\hat{\alpha}_i = \frac{1}{T - T_0}\sum_{t=T_0+1}^{T} \hat{\alpha}_{it} \tag{3.5}$$

for outcome i.

Inference is based on a two-tailed hypothesis test against the null of $\hat{\alpha}_i = 0$. The standard error of $\hat{\alpha}_i$ can be estimated with survey methods. In order to help establish the validity of the test statistic derived from those standard errors, Robbins et al. (2017) propose a permutation test in the spirit of the placebo analysis of Abadie et al. (2010).

The permutation test is structured as follows. First, K total placebo groups are constructed. Each group is selected from the original J observations by randomly re-ordering the observations, taking the last $J - J_0$ units in the re-ordered data as the new placebo treatment group, and the first J_0 units as the new donor pool. New weights, $\{w_j\}^{(k)}$, satisfying versions of the constraints in equations (3.1) through (3.3) are calculated for each placebo group, where the superscript k enumerates the groups. The weights are used to compute a set of K test statistics, $(Z_{i1}, \ldots, Z_{iK})$ for each outcome. Finally a permuted p-value is derived as follows:

$$p_i = \frac{\#k : Z_i^{(k)} < Z_i}{K}. \tag{3.6}$$

Equation (3.6) is expected to hold for sufficiently large K.

3.5 Results

Table 3.1 reports descriptive statistics for the treated group and the full sample, scaled for comparison, during the pre-treatment period. The full sample is on average slightly older than the treated population with comparable distributions of men to women and rural to metropolitan areas. The treated population exhibits a higher distribution of co-morbid conditions indicating greater disease burden among the treated. In addition, the treated group utilize emergency room, primary care and hospital inpatient services at significantly higher rates during the pre-treatment period.

In order to match the treated groups pattern of pre-treatment utilization, I employ the synthetic control method described in Section 4. Following the advice of Ferman et al. (2020), I report results on eight different sets of synthetic controls to avoid specification searching. In my first specification, I match the treated group with a synthetic control group based on emergency room and primary care utilization during the 18-month period prior to treatment with controls for hospital utilization over the same period as well as age, gender, rurality and co-morbidity as of the month prior to treatment. In specification two, I remove the time-invariant constraints. In specification three, I re-introduce the time-invariant constraints and instead eliminate the constraint on hospital utilization. Finally, in specification four, I match treatment and synthetic control groups based on emergency room and primary care utilization only. I then repeat each of these specifications with a shorter 12-month pre-treatment matching period in specifications five through eight.

In all eight specifications I am able to match a synthetic control group to the treated groups' pattern of pre-treatment emergency room and primary care utilization. An exact match is possible because of the aggregate nature of the estimator and the large donor pool. The results of my synthetic control analysis are reported in Table 3.4.

The average treatment effect on the treated for primary care utilization is statistically significant across all eight synthetic control specifications. The estimates range from an increase in primary care utilization of 12.2 percent in specification (7) to 17.1 percent in specification (6). In addition to the consistency of the point estimates, the confidence intervals exclude a null effect and are stable across all eight specifications.

The results on emergency room utilization are less clear cut. The average treatment effect on the treated for emergency room utilization is not statistically significant at conventional levels and is of the wrong sign in six of the eight synthetic control specifications. Furthermore, even when the point estimates are significant, the permuted confidence intervals are very wide and include zero. The wider confidence intervals and higher point estimates appear to be correlated with the inclusion of the time-invariant controls, e.g. compare specification (1) against (2), (3) against (4) and so on.

To further illustrate the increased volatility of specification (1) relative to specification (2), I plot aggregate emergency room and primary care utilization for specification (1) in Figures 3.1 and 3.3 and for specification (2) in Figures 3.2 and 3.4. In the top panel of each figure aggregate utilization for the treated group (solid red line) is plotted against the synthetic control group (dashed black line) and the full scaled sample (dotted green line). The bottom panel plots the difference between the treatment and synthetic control groups aggregate outcomes (solid red line) and differences between placebo treatment and control groups (solid gray lines).

Each figure shows that the synthetic control group matches the treated groups pattern of pre-treatment utilization in both specifications. Comparing Figures

3.3 against 3.4 and Figure 3.1 against 3.2 it is evident that the synthetic groups post-treatment utilization is more volatile when the time-invariant controls are included.

It is also clear from Figures 3.3 and 3.4 that nurse outreach had a positive effect on primary care utilization. With the exception of one period, the treated group accessed primary care at higher rates than the synthetic control and placebo groups throughout the post-treatment period. By contrast the direction of the effect on emergency room utilization depicted in Figures 3.1 and 3.2 is unclear and is plausibly a placebo effect.

3.5.1 Robustness

For comparison, I estimate the relationship between nurse outreach and utilization with a standard two-way fixed effects regression model:

$$y_{it} = D_{it} + x_{it} + \lambda_i + \gamma_t + \epsilon_{it}, \tag{3.7}$$

where y_{it} is the rate of utilization for person i in month t, D_{it} is a treatment indicator, x_{it} is the age of individual i in month t, λ_i is a vector of individual fixed effects and γ_t is a vector of month fixed effects. Heteroscedasticity robust standard errors are used for significance testing.

I report my findings in Table 3.2. The results indicate that treatment reduced emergency room visits by 0.147 visits per person per month and increased primary care visits by 0.070 visits per person per month. Contrary, to the synthetic control analysis, both effects are highly statistically significant and of the expected sign.

However, the differences in the pre-treatment patterns of utilization exhibited by the treatment group and the full sample in Table 3.1 suggest a potential violation of the parallel trends assumption. If the parallel trends assumption is violated, the untreated sample will not provide a suitable counterfactual.

In order to test the parallel trends assumption I conduct an event study. Results from the event study with five trimmed treatment month leads and lags are reported in Table 3.3. If the parallel trends assumption holds, the pre-treatment indicators should not predict outcomes. As expected, the parallel trends assumption is violated. All of the pre-treatment indicators for emergency room utilization and three of the five pre-treatment indicators for primary care utilization are statistically significant at conventional levels indicating a failure of the parallel trends assumption.

3.6 Discussion

This study provides evidence on the impact of a nurse outreach program for Medicare Supplement beneficiaries with a history of frequent emergency department utilization. The program sought to encourage primary care engagement. Primary care engagement has been linked to reduced emergency department utilization and lower costs.

I analyze the impact of nurse outreach using synthetic control methods. To my knowledge, this is the first time that synthetic control methods have been applied to study this type of intervention. I find that nurse outreach did increase primary care engagement. In the six months following treatment, treated individuals utilized primary care 15 percent more often than the synthetic control group.

However, I do not find clear evidence for an effect on emergency department utilization. Prior authors have suggested that the appropriateness of care, intensity of treatment and risk of future utilization are important predictors of a program's ability to reduce emergency room use. Nurse outreach may not have worked because it did not adequately address any one of these criteria. First, it is possible that connecting frequent health care users with existing resources did not adequately address patient needs (Finkelstein et al., 2020). The target population already utilized primary care at a higher rate than the general populace, indicating an increased level of primary care engagement prior to treatment. Primary care engagement may have diminishing returns, particularly for individuals with complex care needs. Second, nurse outreach is a low intensity intervention. The target population may simply need more assistance. Third, the target population in this study experienced a sharp increase followed by a gentle decline in utilization prior to treatment. This pattern of utilization could indicate a degree of episodic rather than chronic illness that would blunt the effectiveness of treatment and make it more difficult to detect an effect.

Table 3.1
Summary statistics in the pre-intervention period

Measure	Treatment (N=895)	Full sample (N=5,445)
Mean age	73.2	76.0
Proportion female	0.59	0.60
Proportion living in a metropolitan area	0.80	0.80
Charleson co-morbidity index (CCI = 1)	120	181.5
Charleson co-morbidity index (CCI = 2)	143	136.9
Charleson co-morbidity index (CCI = 3)	581	279.3
ER visits (t=18)	167	69.7
ER visits (t=17)	233	77.7
ER visits (t=16)	298	90.1
ER visits (t=15)	1,166	229.6
ER visits (t=14)	639	137.6
ER visits (t=13)	403	98.8
ER visits (t=12)	549	121.5
ER visits (t=11)	328	84.7
ER visits (t=10)	318	84.8
ER visits (t=9)	260	72.2
ER visits (t=8)	231	66.4
ER visits (t=7)	182	65.1
ER visits (t=6)	148	51.3
ER visits (t=5)	60	42.4
ER visits (t=4)	64	43.7
ER visits (t=3)	137	51.9
ER visits (t=2)	97	47.5
ER visits (t=1)	95	47.5
Primary care visits (t=18)	1,429	822.7
Primary care visits (t=17)	1,583	858.3
Primary care visits (t=16)	1,537	839.6
Primary care visits (t=15)	1,693	869.9
Primary care visits (t=14)	1,613	850.0
Primary care visits (t=13)	1,523	828.4
Primary care visits (t=12)	1,530	835.3
Primary care visits (t=11)	1,407	814.8
Primary care visits (t=10)	1,442	813.09
Primary care visits (t=9)	1,368	818.7
Primary care visits (t=8)	1,320	803.8
Primary care visits (t=7)	1,308	774.7
Primary care visits (t=6)	1,195	796.2
Primary care visits (t=5)	1,184	779.3
Primary care visits (t=4)	1,195	775.8
Primary care visits (t=3)	1,188	767.8
Primary care visits (t=2)	1,195	767.3
Primary care visits (t=1)	1,118	759.9
Hospital days (t $\geq$ 18 and t $<$ 12)	4,178	1,097
Hospital days (t $\geq$ 12 and t $<$ 6)	3,112	858
Hospital days (t $\geq$ 6 and t $<$ 0)	1,215	538

Data obtained from demographic and commercial claims data as part of a data use agreement with a private Medicare Supplement insurance provider. Measures correspond to the $1 + L + IT_0 = 46$ constraints outlined in equations (3.1) through (3.3). The unit of analysis is an individual-month. The Full Sample is scaled to be proportional to the treated group.

Table 3.2
Two-way fixed effect estimates of the effect of nurse outreach on health care utilization

	Emergency room visits	Primary care visits
Treat·Post	-0.147^{***}	0.070^{***}
	(0.012)	(0.027)
N	5,445	5,445
Mean	0.069	0.947

Effect of nurse outreach on emergency room and primary care visits. The unit of analysis is an individual-month. The model includes control variables and individual and month fixed effects. Cluster robust standard errors are in parentheses. Statistical significance: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Table 3.3
Two-way fixed effect event study

Lead-/Lag+	Emergency room visits	Primary care visits
-5	0.465^{***}	0.337^{***}
	(0.044)	(0.052)
-4	0.582^{***}	0.080
	(0.080)	(0.065)
-3	-0.974^{***}	-0.168^{***}
	(0.071)	(0.062)
-2	-0.070^{*}	0.036
	(0.041)	(0.063)
-1	-0.077^{**}	-0.158^{***}
	(0.027)	(0.060)
0	-0.004	0.137^{**}
	(0.021)	(0.062)
1	-0.013	-0.116^{**}
	(0.022)	(0.059)
2	0.054^{**}	-0.051
	(0.027)	(0.056)
3	-0.039	0.078
	(0.025)	(0.059)
4	0.031	-0.076
	(0.032)	(0.057)
5	0.000	0.004
	(0.034)	(0.056)
N	5,445	5,445
Mean	0.069	0.947

Effect of nurse outreach on emergency room and primary care visits. The unit of analysis is an individual-month. The model includes control variables and individual and month fixed effects. Cluster robust standard errors are in parentheses. Statistical significance: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

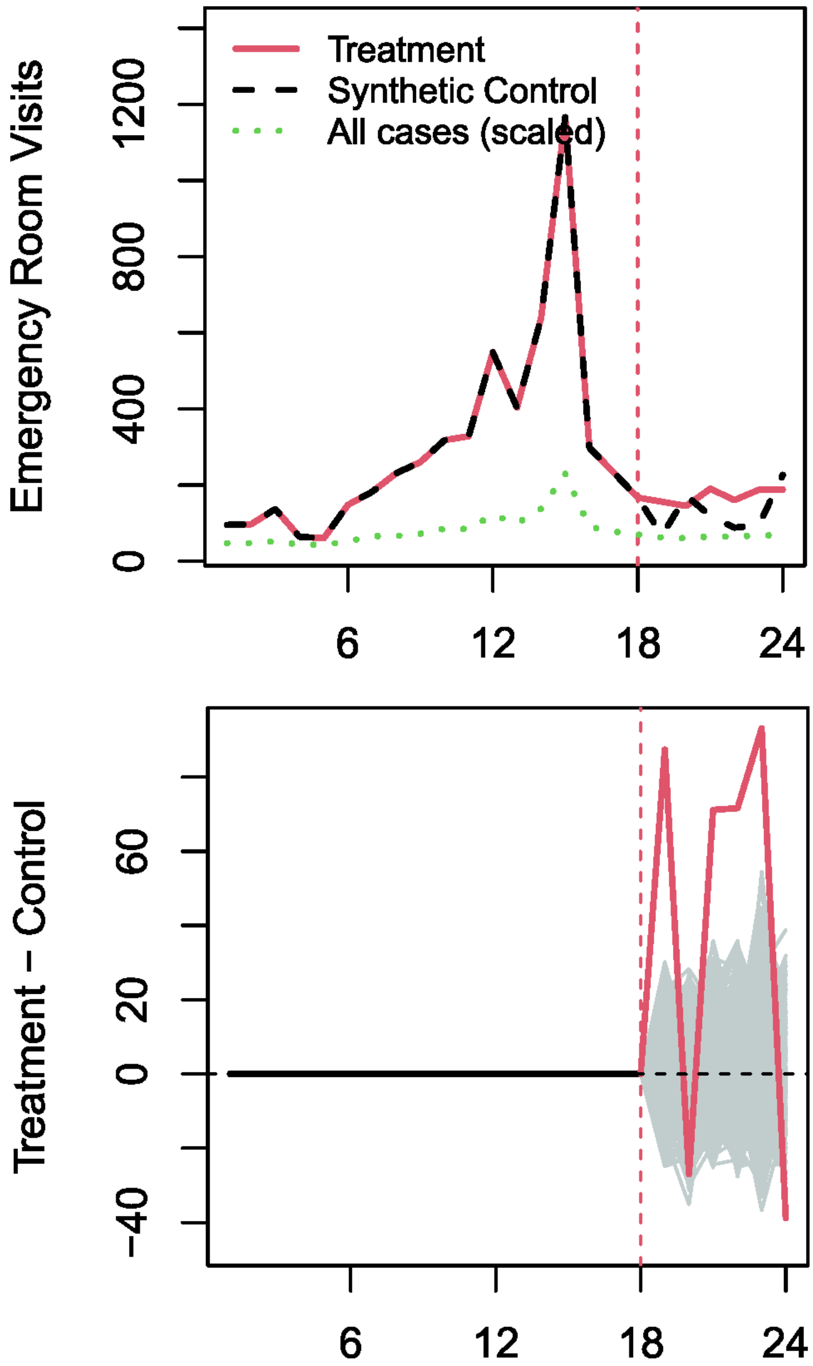
Table 3.4

Synthetic control estimates of the effect of nurse outreach on health care utilization

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Pre-treatment:								
18-months	Yes	Yes	Yes	Yes	No	No	No	No
12-months	No	No	No	No	Yes	Yes	Yes	Yes
Hospital days	Yes	Yes	No	No	Yes	Yes	No	No
Controls	Yes	No	Yes	No	Yes	No	Yes	No
ER visits %Δ								
Treat - SC	33.5**	10.0	29.2*	10.2	9.1	-1.3	6.4	-2.5
95% upper CI	80.6	41.0	87.7	38.4	53.3	27.6	56.1	27.1
95% lower CI	-2.8	-11.2	-6.5	-14.0	-19.8	-23.4	-25.3	-23.1
PCP visits %Δ								
Treat - SC	15.4**	15.7***	13.8**	15.1***	12.9**	17.1***	12.2**	16.0***
95% upper CI	29.2	26.1	26.7	24.2	25.2	26.0	23.7	26.1
95% lower CI	3.1	6.5	3.0	6.1	3.3	9.0	2.1	7.4
N	5,445	5,445	5,445	5,445	5,445	5,445	5,445	5,445

Effect of nurse outreach on emergency room and primary care utilization. The unit of analysis is an individual-month. Model (1) includes all constraints in Table 3.1. Model (2) excludes the constraints on age, gender, rurality and CCI. Model (3) excludes the constraints on hospital utilization. Model (4) excludes the constraints on age, gender, rurality, CCI and hosptial utilization. Models (5) - (8) mirror the constraints in models (1) - (4), but limit the pre-treatment matching period to the twelve months prior to treatment. p-values and confidence intervals are derived from $B = 500$ placebo tests. See Robbins et al. (2017) for details. Statistical significance: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

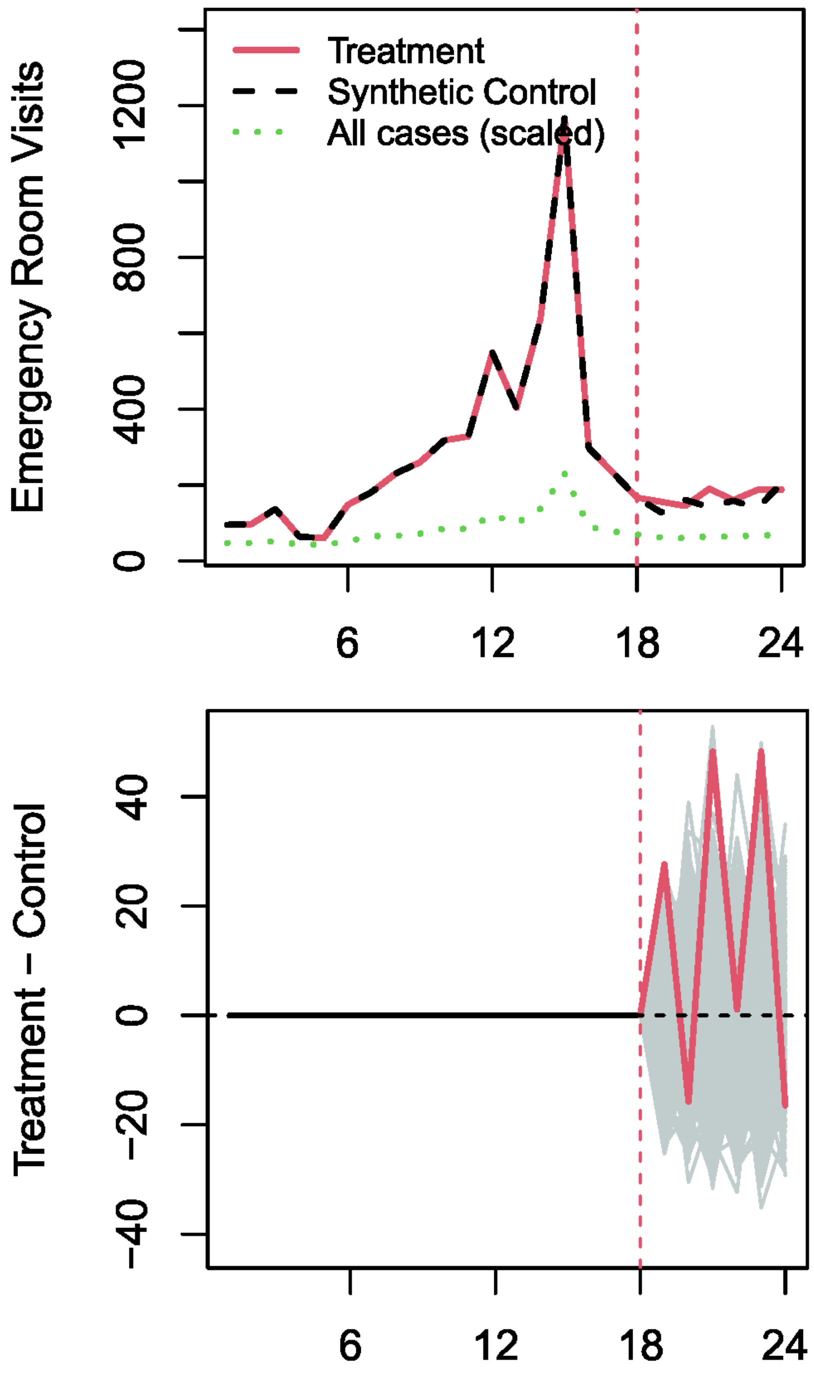
Figure 3.1
Emergency room visits over time: Treatment vs Synthetic Control - Model (1)



Effect of nurse outreach on emergency room utilization based on Model (1). The top panel plots aggregate emergency room uitilization over time for the treated group (red solid line), synthetic control (black dashed line) and the full (scaled) sample (dotted green line). The bottom panel plots the aggregate difference between the treated (solid red line) and synthetic control groups (black line). Differences between placebo groups are plotted in gray. The vertical dashed line indicates the end of the pre-intervention period.

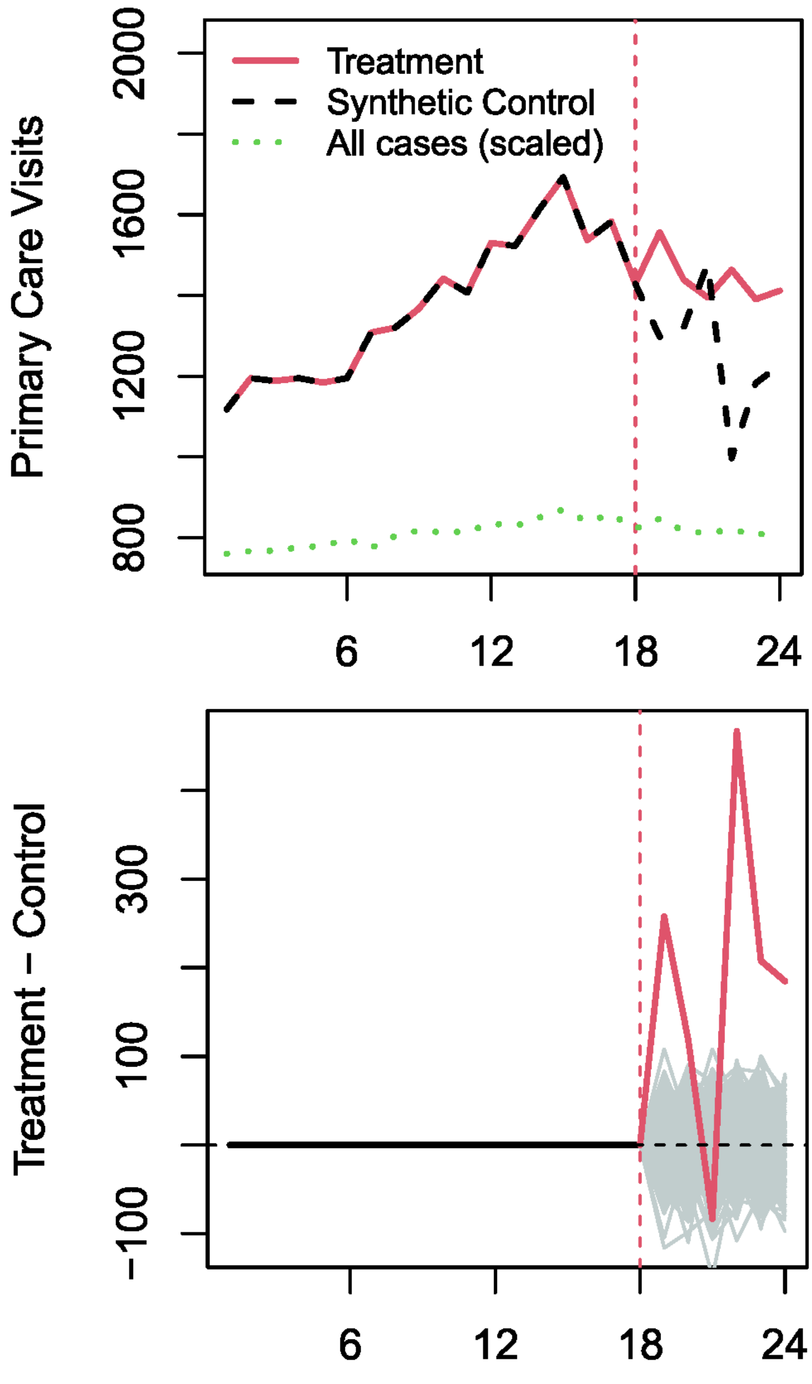
Figure 3.2

Emergency room visits over time: Treatment vs Synthetic Control - Model (2)



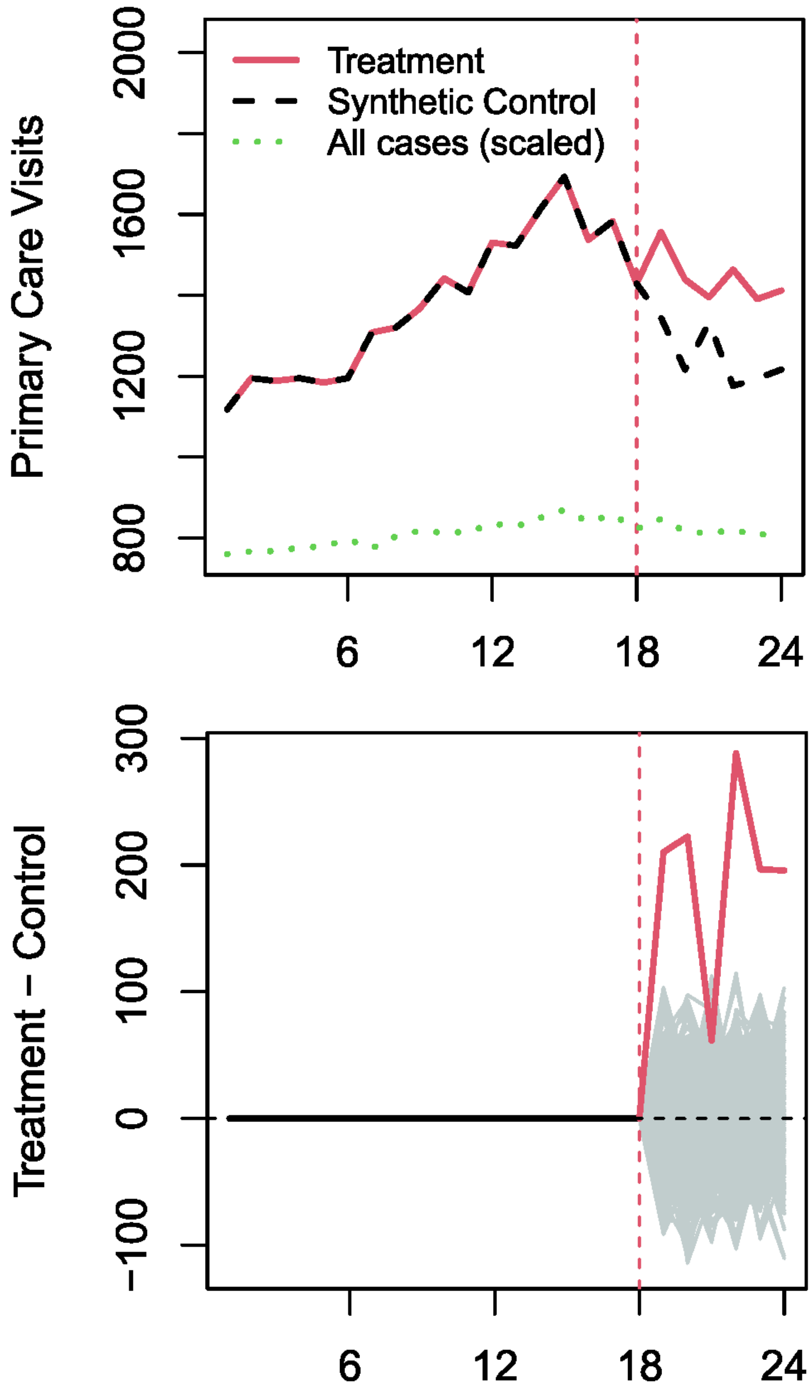
Effect of nurse outreach on emergency room utilization based on Model (2). The top panel plots aggregate emergency room uitilization over time for the treated group (red solid line), synthetic control (black dashed line) and the full (scaled) sample (dotted green line). The bottom panel plots the aggregate difference between the treated (solid red line) and synthetic control groups (black line). Differences between placebo groups are plotted in gray. The vertical dashed line indicates the end of the pre-intervention period.

Figure 3.3
Primary care visits over time: Treatment vs Synthetic Control - Model (1)



Effect of nurse outreach on primary care utilization based on Model (1). The top panel plots aggregate primary care uitilization over time for the treated group (red solid line), synthetic control (black dashed line) and the full (scaled) sample (dotted green line). The bottom panel plots the aggregate difference between the treated (solid red line) and synthetic control groups (black line). Differences between placebo groups are plotted in gray. The vertical dashed line indicates the end of the pre-intervention period.

Figure 3.4
Primary care visits over time: Treatment vs Synthetic Control - Model (2)



Effect of nurse outreach on primary care utilization based on Model (2). The top panel plots aggregate primary care uitilization over time for the treated group (red solid line), synthetic control (black dashed line) and the full (scaled) sample (dotted green line). The bottom panel plots the aggregate difference between the treated (solid red line) and synthetic control groups (black line). Differences between placebo groups are plotted in gray. The vertical dashed line indicates the end of the pre-intervention period.

Bibliography

Abadie, A., Diamond, A. and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California‘s tobacco control program, *Journal of the American Statistical Association* **105**(490): 493–505.

Abadie, A. and LHour, J. (2021). A penalized synthetic control estimator for disaggregated data, *Journal of the American Statistical Association* **116**(536): 1817–1834.

Advani, A. and Malde, B. (2018). Credibly identifying social effects: Accounting for network formation and measurement error, *Journal of Economic Surveys* **32**(4): 1016–1044.

Althaus, F., Paroz, S., Hugli, O., Ghali, W. A., Daeppen, J.-B., Peytremann-Bridevaux, I. and Bodenmann, P. (2011). Effectiveness of interventions targeting frequent users of emergency departments: A systematic review, *Annals of Emergency Medicine* **58**(1): 41–52.

Ammermueller, A. and Pischke, J.-S. (2009). Peer effects in European primary schools: Evidence from the Progress in International Reading Literacy Study, *Journal of Labor Economics* **27**(3): 315–348.

Banerjee, A., Breza, E., Chandrasekhar, A. G., Duflo, E., Jackson, M. O. and Kinnan, C. (2021). Changes in social network structure in response to exposure to formal credit markets, *Technical report*, National Bureau of Economic Research.

Banerjee, A., Chandrasekhar, A. G., Duflo, E. and Jackson, M. O. (2013). The diffusion of microfinance, *Science* **341**(6144).

Blewett, L. A., Johnson, P. J., Lee, B. and Scal, P. B. (2008). When a usual source of care and usual provider matter: Adult prevention and screening services, *Journal of general internal medicine* **23**(9): 1354–1360.

Blume, L. E., Brock, W. A., Durlauf, S. N. and Jayaraman, R. (2015). Linear social interactions models, *Journal of Political Economy* **123**(2): 444–496.

Bramoullé, Y., Djebbari, H. and Fortin, B. (2009). Identification of peer effects through social networks, *Journal of Econometrics* **150**(1): 41–55.

Bramoullé, Y., Djebbari, H. and Fortin, B. (2020). Peer effects in networks: A survey, *Annual Review of Economics* **12**: 603–629.

Burns, T., Catty, J., Dash, M., Roberts, C., Lockwood, A. and Marshall, M. (2007). Use of intensive case management to reduce time in hospital in people with severe mental illness: Systematic review and meta-regression, *BMJ* **335**(7615): 336.

Cai, J., Janvry, A. D. and Sadoulet, E. (2015). Social networks and the decision to insure, *American Economic Journal: Applied Economics* **7**(2): 81–108.

Carrell, S. E., Figlio, D. N. and Lusher, L. R. (2022). Clubs and Networks in Economics Reviewing, *Technical report*, National Bureau of Economic Research.

Chandrasekhar, A. and Lewis, R. (2011). Econometrics of sampled networks, *Unpublished manuscript, MIT.[422]* .

Choudhry, N. K., Fischer, M. A., Avorn, J. L., Lee, J. L., Schneeweiss, S., Solomon, D. H., Berman, C., Jan, S., Lii, J., Mahoney, J. J. et al. (2012). The impact of reducing cardiovascular medication copayments on health spending and resource utilization, *Journal of the American College of Cardiology* **60**(18): 1817–1824.

De Giorgi, G., Pellizzari, M. and Redaelli, S. (2010). Identification of social interactions through partially overlapping peer groups, *American Economic Journal: Applied Economics* **2**(2): 241–75.

D'Hoore, W., Bouckaert, A. and Tilquin, C. (1996). Practical considerations on the use of the Charlson comorbidity index with administrative data bases, *Journal of Clinical Epidemiology* **49**(12): 1429–1433.

Ferman, B., Pinto, C. and Possebom, V. (2020). Cherry picking with synthetic controls, *Journal of Policy Analysis and Management* **39**(2): 510–532.

Finkelstein, A., Zhou, A., Taubman, S. and Doyle, J. (2020). Health care hotspottinga randomized, controlled trial, *New England Journal of Medicine* **382**(2): 152–162.

Freeman, H. P. (2006). Patient navigation: A community based strategy to reduce cancer disparities, *Journal of Urban Health* **83**(2): 139–141.

Griffith, A. (2019). Name your friends, but only five? The importance of censoring in peer effects estimates using social network data.

Halliday, T. J. and Kwak, S. (2012). What is a peer? The role of network definitions in estimation of endogenous peer effects, *Applied Economics* **44**(3): 289–302.

Hirth, R. A., Cliff, E. Q., Gibson, T. B., McKellar, M. R. and Fendrick, A. M. (2016). Connecticuts value-based insurance plan increased the use of targeted services and medication adherence, *Health Affairs* **35**(4): 637–646.

Hsu, J., Price, M., Brand, R., Ray, G. T., Fireman, B., Newhouse, J. P. and Selby, J. V. (2006). Cost-sharing for emergency care and unfavorable clinical events: Findings from the Safety and Financial Ramifications of ed Copayments Study, *Health Services Research* **41**(5): 1801–1820.

Hudon, C., Chouinard, M.-C., Pluye, P., El Sherif, R., Bush, P. L., Rihoux, B., Poitras, M.-E., Lambert, M., Zomahoun, H. T. V. and Légaré, F. (2019). Characteristics of case management in primary care associated with positive outcomes for frequent users of health care: A systematic review, *The Annals of Family Medicine* **17**(5): 448–458.

Insinga, R. P., Ng-Mak, D. S. and Hanson, M. E. (2011). Costs associated with outpatient, emergency room and inpatient care for migraine in the USA, *Cephalalgia* **31**(15): 1570–1575.

Kandpal, E. and Baylis, K. (2013). Expanding horizons: Can women's support groups diversify peer networks in rural India?, *American Journal of Agricultural Economics* **95**(2): 360–367.

Keehan, S. P., Stone, D. A., Poisal, J. A., Cuckler, G. A., Sisko, A. M., Smith, S. D., Madison, A. J., Wolfe, C. J. and Lizonitz, J. M. (2017). National health expenditure projections, 2016–25: Price increases, aging push sector to 20 percent of economy, *Health Affairs* **36**(3): 553–563.

Kumar, G. S. and Klein, R. (2013). Effectiveness of case management strategies in reducing emergency department visits in frequent user patient populations: A systematic review, *The Journal of Emergency Medicine* **44**(3): 717–729.

Laschever, R. (2011). The doughboys network: Social interactions and labor market outcomes of World War I veterans, *Unpublished manuscript* .

Liu, X. (2013). Estimation of a local-aggregate network model with sampled networks, *Economics Letters* **118**(1): 243–246.

Liu, X. and Lee, L.-f. (2010). GMM estimation of social interaction models with centrality, *Journal of Econometrics* **159**(1): 99–115.

Liu, X., Patacchini, E. and Zenou, Y. (2014). Endogenous peer effects: Local aggregate or local average?, *Journal of Economic Behavior & Organization* **103**: 39–59.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem, *The Review of Economic Studies* **60**(3): 531–542.

McGinnis, J. M., Stuckhardt, L., Saunders, R., Smith, M. et al. (2013). *Best Care at Lower Cost: The Path to Continuously Learning Health Care in America*, National Academies Press.

Moore, B. J. and Liang, L. (2020). *Costs of emergency department visits in the United States, 2017*, Agency for Healthcare Research and Quality.

Nair, K. V., Miller, K., Park, J., Allen, R. R., Saseen, J. J. and Biddle, V. (2010). Prescription co-pay reduction program for diabetic employees, *Population Health Management* **13**(5): 235–245.

Navratil-Strawn, J. L., Hawkins, K., Wells, T. S., Ozminkowski, R. J., Hartley, S. K., Migliori, R. J. and Yeh, C. S. (2014). An emergency room decision-support program that increased physician office visits, decreased emergency room visits, and saved money, *Population Health Management* **17**(5): 257–264.

Newman, M. E. and Ziff, R. M. (2001). Fast Monte Carlo algorithm for site or bond percolation, *Physical Review E* **64**(1): 016706.

Newton, M. F., Keirns, C. C., Cunningham, R., Hayward, R. A. and Stanley, R. (2008). Uninsured adults presenting to US emergency departments: Assumptions vs data, *JAMA* **300**(16): 1914–1924.

Norris, S. L., Nichols, P. J., Caspersen, C. J., Glasgow, R. E., Engelgau, M. M., Jack Jr, L., Isham, G., Snyder, S. R., Carande-Kulis, V. G., Garfield, S. et al. (2002). The effectiveness of disease and case management for people with diabetes: A systematic review, *American Journal of Preventive Medicine* **22**(4): 15–38.

Oster, E. and Thornton, R. (2012). Determinants of technology adoption: Peer effects in menstrual cup take-up, *Journal of the European Economic Association* **10**(6): 1263–1293.

Patacchini, E., Rainone, E. and Zenou, Y. (2017). Heterogeneous peer effects in education, *Journal of Economic Behavior & Organization* **134**: 190–227.

Robbins, M. W., Saunders, J. and Kilmer, B. (2017). A framework for synthetic control methods with high-dimensional, micro-level data: Evaluating a neighborhood-specific crime intervention, *Journal of the American Statistical Association* **112**(517): 109–126.

Rosenblatt, R. A., Wright, G. E., Baldwin, L.-M., Chan, L., Clitherow, P., Chen, F. M. and Hart, L. G. (2000). The effect of the doctor-patient relationship on emergency department use among the elderly, *American Journal of Public Health* **90**(1).

Sabety, A. H., Jena, A. B. and Barnett, M. L. (2021). Changes in health care use and outcomes after turnover in primary care, *JAMA internal medicine* **181**(2): 186–194.

Siddiqui, M., Roberts, E. T. and Pollack, C. E. (2015). The effect of emergency department copayments for Medicaid beneficiaries following the Deficit Reduction Act of 2005, *JAMA Internal Medicine* **175**(3): 393–398.

Sojourner, A. (2013). Identification of peer effects with missing peer data: Evidence from Project STAR, *The Economic Journal* **123**(569): 574–605.

Soril, L. J., Leggett, L. E., Lorenzetti, D. L., Noseworthy, T. W. and Clement, F. M. (2015). Reducing frequent visits to the emergency department: A systematic review of interventions, *PloS One* **10**(4).

The Board of Trustees, F. H. I. and Funds, F. S. M. I. T. (2021). 2021 Annual Report of the Boards of Trustees of the Federal Hospital Insurance and Federal Supplementary Medical Insurance Trust Funds, *Technical report*.

Van den Heede, K. and Van de Voorde, C. (2016). Interventions to reduce emergency department utilisation: A review of reviews, *Health Policy* **120**(12): 1337–1349.

Wells, K. J., Battaglia, T. A., Dudley, D. J., Garcia, R., Greene, A., Calhoun, E., Mandelblatt, J. S., Paskett, E. D., Raich, P. C. and Program, P. N. R. (2008). Patient navigation: State of the art or is it science?, *Cancer* **113**(8): 1999–2010.

Ziguras, S. J. and Stuart, G. W. (2000). A meta-analysis of the effectiveness of mental health case management over 20 years, *Psychiatric Services* **51**(11): 1410–1421.