Causal Inference for Infectious Disease Prevention

Brian Richardson

Department of Biostatistics,
University of North Carolina at Chapel Hill

November 17, 2025

Scan for slides

Outline

- About Me

Outline

- About Me
- Intro to Biostatistics

Outline

- About Me
- Intro to Biostatistics
- Intro to Causal Inference

Outline

- About Me
- Intro to Biostatistics
- Intro to Causal Inference
- Research Project: Social Distancing to Reduce Transmission of the Flu on College Campuses

About Me

About Me

About Me


Richmond, VA
Greensboro
Chapel Hill
Durham
Raleigh
CARRBORO
Asheville
Charlotte
Greenville
New Bern
Greenville, SC
Wilmington

About Me

About Me

About Me

UNIVERSITY OF NORTH CAROLINA CHAPEL HILL LVX LIBERTAS

UNC CFAR Center for AIDS Research

About Me


QB
ME
Ticonderoga
Bar
Harbor
NY
VT
NH
MA
PA
CT
RI
NJ
MD
DE
VA
NORTHERN TIER
Bicycle Route
Adventure Cycling Association

About Me


QB
ME
Ticonderoga
Bar
Harbor
NY
VT
NH
MA
PA
CT
RI
NJ
MD
DE
VA
NORTHERN TIER
Bicycle Route
Adventure Cycling Association


STOP

About Me


QB
ME
Ticonderoga
Bar
Harbor
NY
VT
NH
MA
PA
CT
RI
NJ
MD
DE
VA
NORTHERN TIER
Bicycle Route
Adventure Cycling Association

About Me


QB
ME
Ticonderoga
Bar
Harbor
NY
VT NH
MA
PA
CT
RI
NJ
MD DE
VA
NORTHERN TIER
Bicycle Route
Adventure Cycling Association

Why Biostatistics?

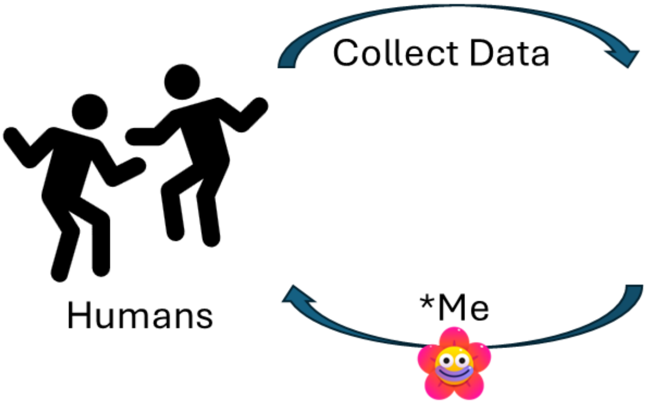
Why Biostatistics?

Humans

Data

X	Y	Z
0	100	H
0	50	L
1	100	H
1	120	H

Why Biostatistics?



Data

X	Y	Z
0	100	H
0	50	L
1	100	H
1	120	H

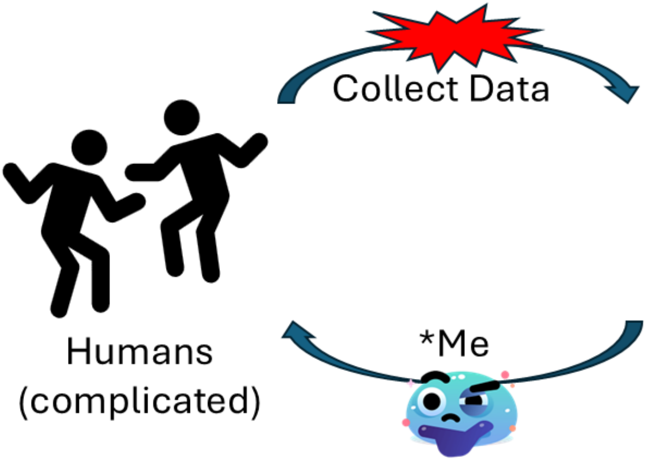
Why Biostatistics?

Humans
(complicated)

Data

X	Y	Z
0	100	H
NA	>100	L or H
NA	100	L or H
1	>100	H

Why Biostatistics?



Data

X	Y	Z
0	100	H
NA	>100	L or H
NA	100	L or H
1	>100	H

Why Biostatistics?

Collect Data

Humans
(complicated)
(and precious)

Data

X	Y	Z
0	100	H
NA	>100	L or H
NA	100	L or H
1	>100	H

*Me

Causal Inference

We are often interested in **causation**, as opposed to **association**

Causal Inference

We are often interested in **causation**, as opposed to **association**

- **Associational question**: Do people who get the COVID-19 vaccine get sick?

Causal Inference

We are often interested in **causation**, as opposed to **association**

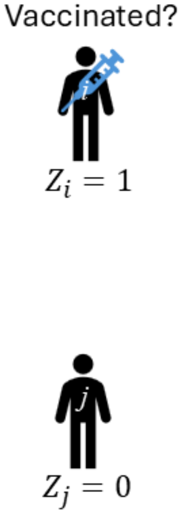
- **Associational question**: Do people who get the COVID-19 vaccine get sick?
- **Causal question**: If someone *were* to get the COVID-19 vaccine, *would* they get sick?

Associational Question

Do people who get the COVID-19 vaccine get sick?

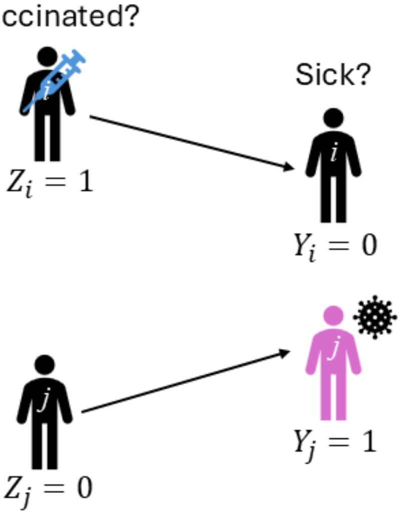
Associational Question

Do people who get the COVID-19 vaccine get sick?



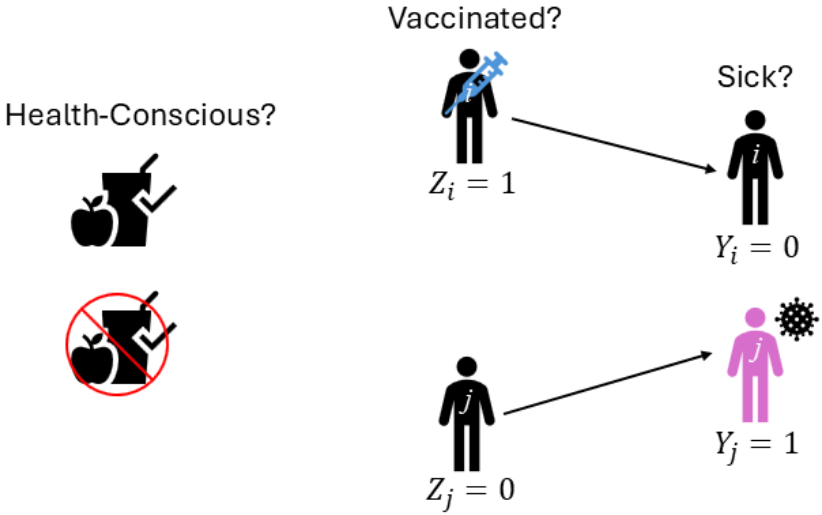
Associational Question

Do people who get the COVID-19 vaccine get sick?



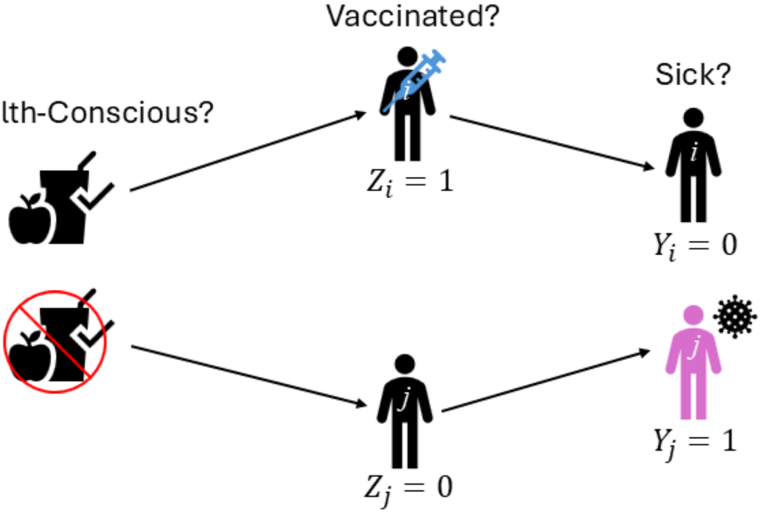
Associational Question

Do people who get the COVID-19 vaccine get sick?



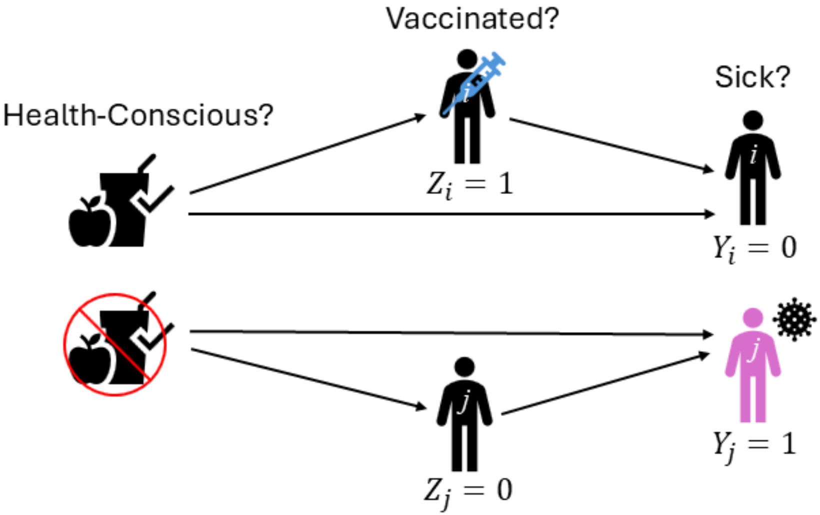
Associational Question

Do people who get the COVID-19 vaccine get sick?



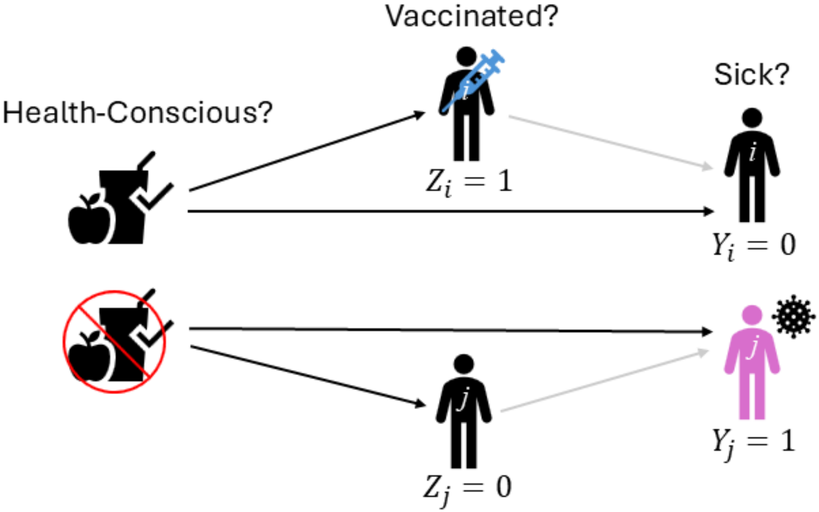
Associational Question

Do people who get the COVID-19 vaccine get sick?



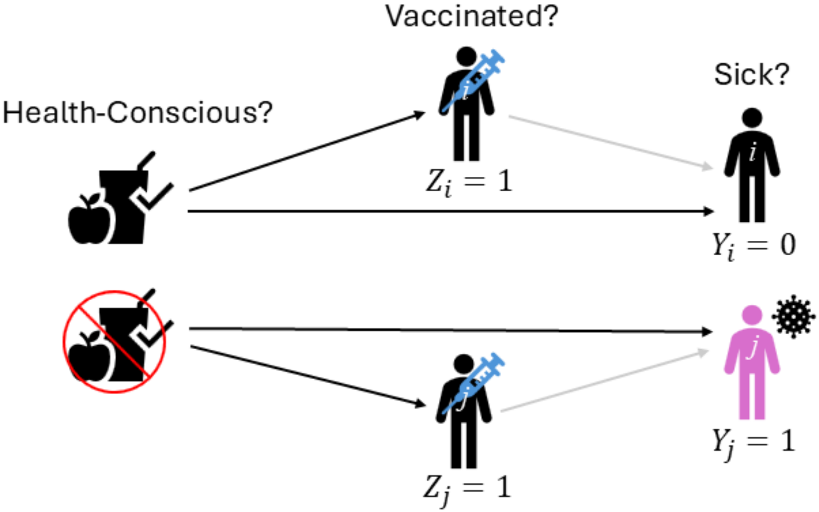
Associational Question

Do people who get the COVID-19 vaccine get sick?



Associational Question

Do people who get the COVID-19 vaccine get sick?

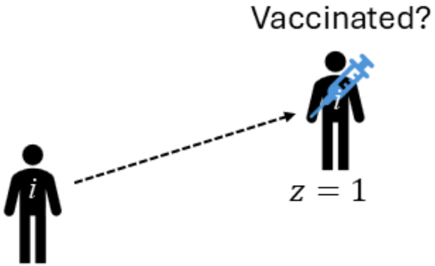


Causal Question

If someone *were* to get the COVID-19 vaccine, *would* they get sick?

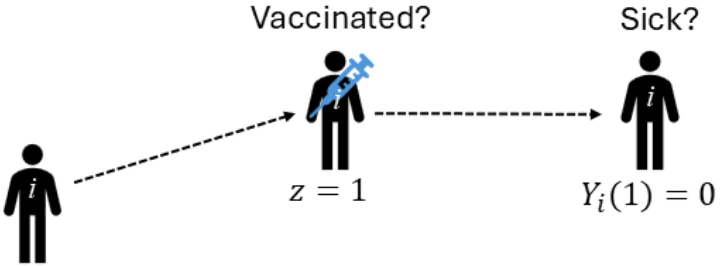
Causal Question

If someone *were* to get the COVID-19 vaccine, *would* they get sick?



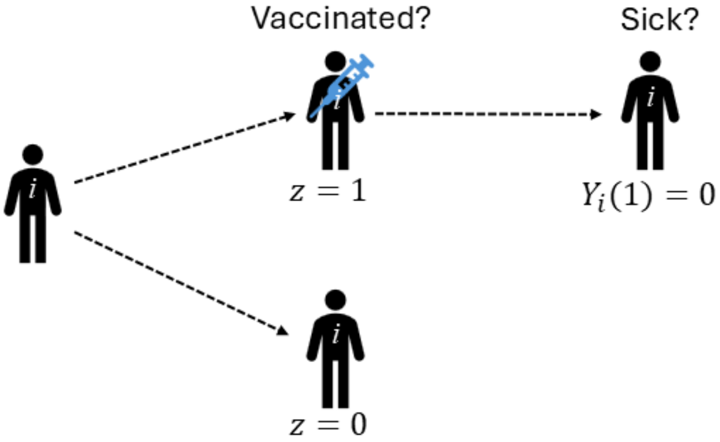
Causal Question

If someone *were* to get the COVID-19 vaccine, *would* they get sick?



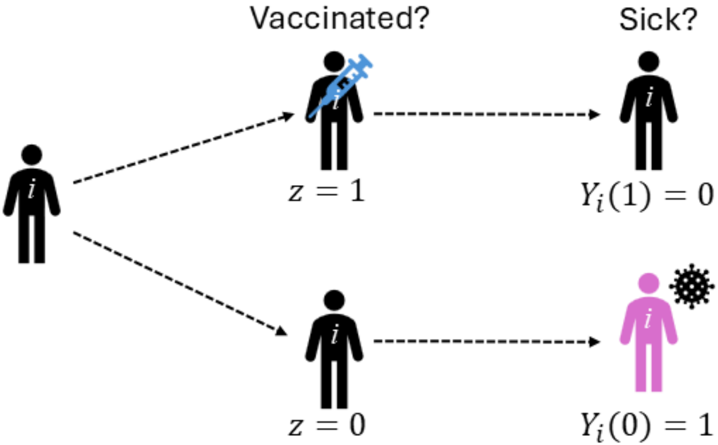
Causal Question

If someone *were* to get the COVID-19 vaccine, *would* they get sick?



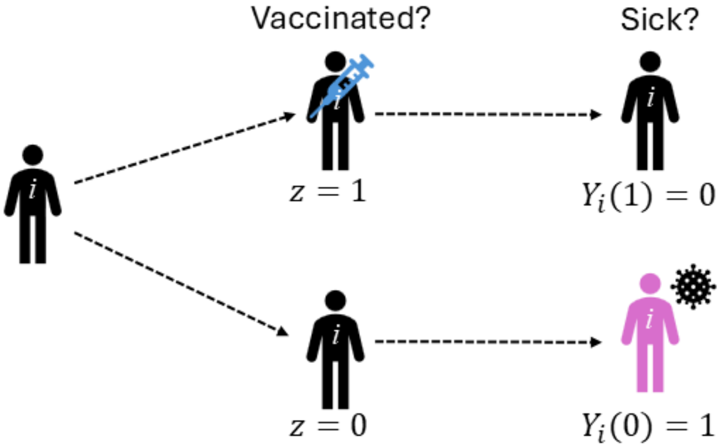
Causal Question

If someone *were* to get the COVID-19 vaccine, *would* they get sick?



Causal Question

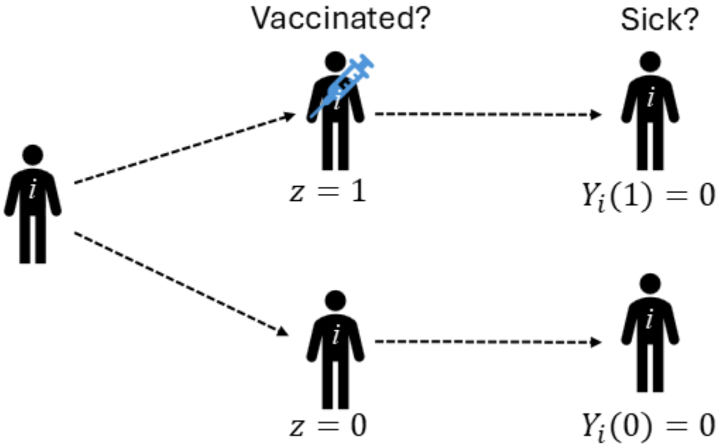
If someone *were* to get the COVID-19 vaccine, *would* they get sick?



$Y_i(1)$	$Y_i(0)$	Type
0	1	Protected

Causal Question

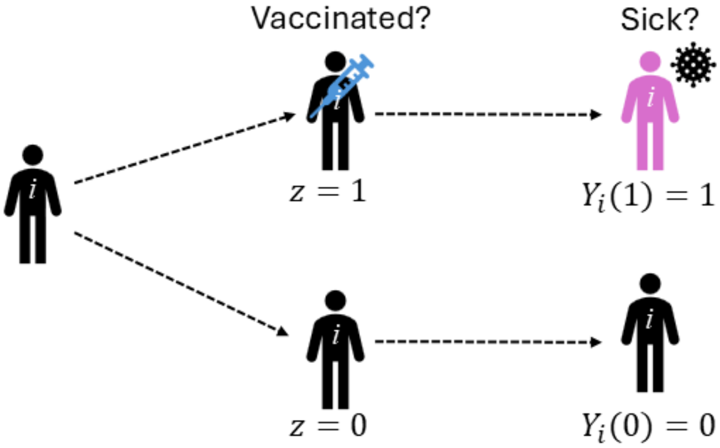
If someone *were* to get the COVID-19 vaccine, *would* they get sick?



$Y_i(1)$	$Y_i(0)$	Type
0	0	Immune
0	1	Protected

Causal Question

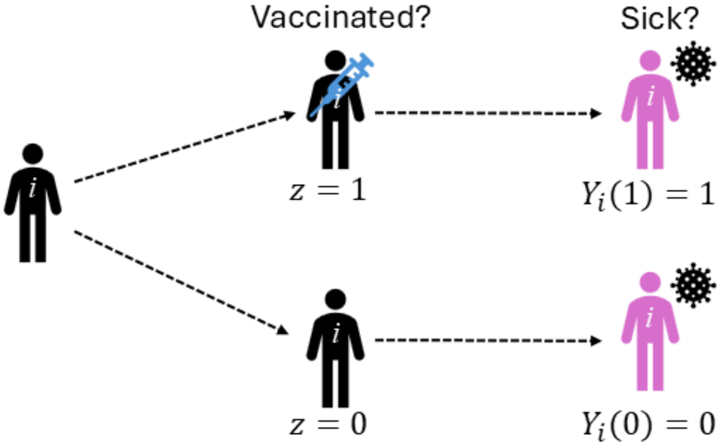
If someone *were* to get the COVID-19 vaccine, *would* they get sick?



$Y_i(1)$	$Y_i(0)$	Type
0	0	Immune
0	1	Protected
1	0	Harmed

Causal Question

If someone *were* to get the COVID-19 vaccine, *would* they get sick?



$Y_i(1)$	$Y_i(0)$	Type
0	0	Immune
0	1	Protected
1	0	Harmed
1	1	Doomed

My Causal Questions

1. If somebody were to have more antibody-dependent cellular phagocytosis (ADCP), would this lower their risk of HIV infection?

My Causal Questions

1. If somebody were to have more antibody-dependent cellular phagocytosis (ADCP), would this lower their risk of HIV infection?

2. If better HIV prevention and treatment resources were available in a community, would this lower the incidence of HIV in that community?

My Causal Questions

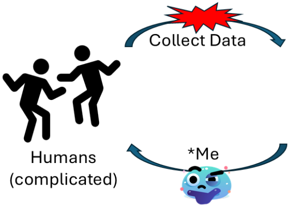
1. If somebody were to have more antibody-dependent cellular phagocytosis (ADCP), would this lower their risk of HIV infection?
2. If better HIV prevention and treatment resources were available in a community, would this lower the incidence of HIV in that community?
3. If college students were told to social distance when sick with the flu, would this reduce transmission of the flu?

My Causal Questions

1. If somebody were to have more antibody-dependent cellular phagocytosis (ADCP), would this lower their risk of HIV infection?

2. If better HIV prevention and treatment resources were available in a community, would this lower the incidence of HIV in that community?

3. If college students were told to social distance when sick with the flu, would this reduce transmission of the flu?

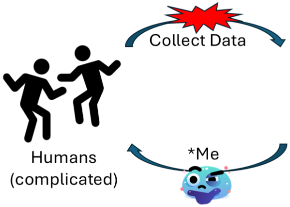


Data

X	Y	Z
0	100	H
NA	>100	L or H
NA	100	L or H
1	>100	H

My Causal Questions

1. If somebody were to have more antibody-dependent cellular phagocytosis (ADCP), would this lower their risk of HIV infection?
   - ADCP is measured with error
2. If better HIV prevention and treatment resources were available in a community, would this lower the incidence of HIV in that community?
3. If college students were told to social distance when sick with the flu, would this reduce transmission of the flu?

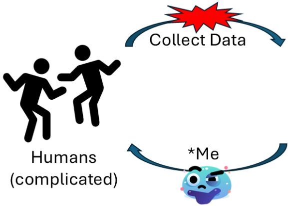


Data

X	Y	Z
0	100	H
NA	>100	L or H
NA	100	L or H
1	>100	H

My Causal Questions

1. If somebody were to have more antibody-dependent cellular phagocytosis (ADCP), would this lower their risk of HIV infection?
   - ADCP is measured with error
2. If better HIV prevention and treatment resources were available in a community, would this lower the incidence of HIV in that community?
   - HIV outcome data are missing (not at random)
3. If college students were told to social distance when sick with the flu, would this reduce transmission of the flu?

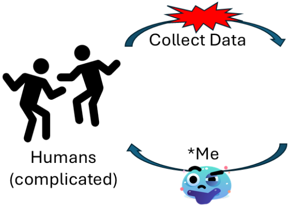


Data

X	Y	Z
0	100	H
NA	>100	L or H
NA	100	L or H
1	>100	H

My Causal Questions

1. If somebody were to have more antibody-dependent cellular phagocytosis (ADCP), would this lower their risk of HIV infection?
   - ADCP is measured with error
2. If better HIV prevention and treatment resources were available in a community, would this lower the incidence of HIV in that community?
   - HIV outcome data are missing (not at random)
3. If college students were told to social distance when sick with the flu, would this reduce transmission of the flu?
   - Network-dependent data, self-reported network



Data

X	Y	Z
0	100	H
NA	**>100**	**L or H**
NA	100	**L or H**
1	**>100**	H

Social Distancing

If college students were told to social distance when sick with the flu, would this reduce transmission of the flu?

eX-FLU Trial

- **eX-FLU**: trial to evaluate a social distancing intervention on a college campus during flu season (Aiello et al., 2016; Zivich et al., 2020)

Design and methods of a social network isolation study for reducing respiratory infection transmission: The eX-FLU cluster randomized trial

Allison E. Aiello[a,*], Amanda M. Simanek[b,1], Marisa C. Eisenberg[c], Alison R. Walsh[c], Brian Davis[c], Erik Volz[d,1], Caroline Cheng[c], Jeanette J. Rainey[e], Amra Uzicanin[e], Hongjiang Gao[e], Nathaniel Osgood[f], Dylan Knowles[f], Kevin Stanley[f], Kara Tarter[c], Arnold S. Monto[c]

eX-FLU Trial

- **eX-FLU**: trial to evaluate a social distancing intervention on a college campus during flu season (Aiello et al., 2016; Zivich et al., 2020)
- **Intervention**: encouragement to isolate in dorm for three days upon developing symptoms of **influenza-like illness** (ILI)

Design and methods of a social network isolation study for reducing respiratory infection transmission: The eX-FLU cluster randomized trial

Allison E. Aiello[a,*], Amanda M. Simanek[b,1], Marisa C. Eisenberg[c], Alison R. Walsh[c], Brian Davis[c], Erik Volz[d,1], Caroline Cheng[c], Jeanette J. Rainey[e], Amra Uzicanin[e], Hongjiang Gao[e], Nathaniel Osgood[f], Dylan Knowles[f], Kevin Stanley[f], Kara Tarter[c], Arnold S. Monto[c]

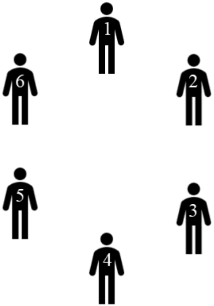
eX-FLU Trial

- **eX-FLU**: trial to evaluate a social distancing intervention on a college campus during flu season (Aiello et al., 2016; Zivich et al., 2020)
- **Intervention**: encouragement to isolate in dorm for three days upon developing symptoms of **influenza-like illness** (ILI)
- **Central question**: does the encouragement-to-isolate intervention reduce transmission of ILI?

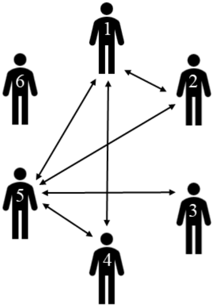
Design and methods of a social network isolation study for reducing respiratory infection transmission: The eX-FLU cluster randomized trial

Allison E. Aiello[a,*], Amanda M. Simanek[b,1], Marisa C. Eisenberg[c], Alison R. Walsh[c], Brian Davis[c], Erik Volz[d,1], Caroline Cheng[c], Jeanette J. Rainey[e], Amra Uzicanin[e], Hongjiang Gao[e], Nathaniel Osgood[f], Dylan Knowles[f], Kevin Stanley[f], Kara Tarter[c], Arnold S. Monto[c]

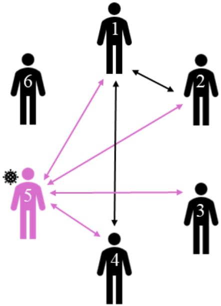
Transmission of Influenza-Like-Illness



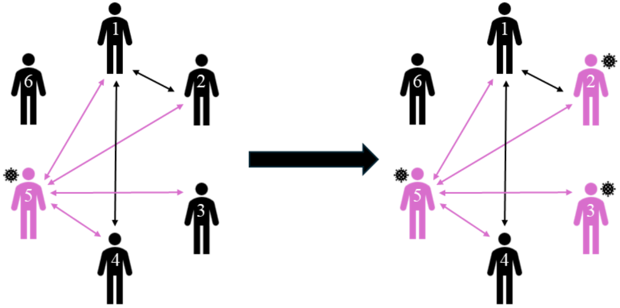
Transmission of Influenza-Like-Illness



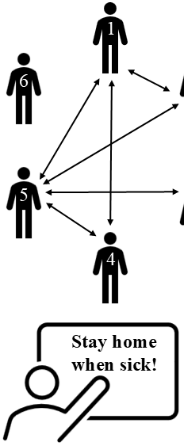
Transmission of Influenza-Like-Illness



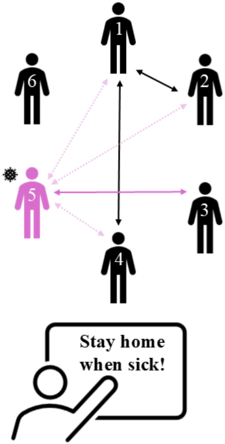
Transmission of Influenza-Like-Illness



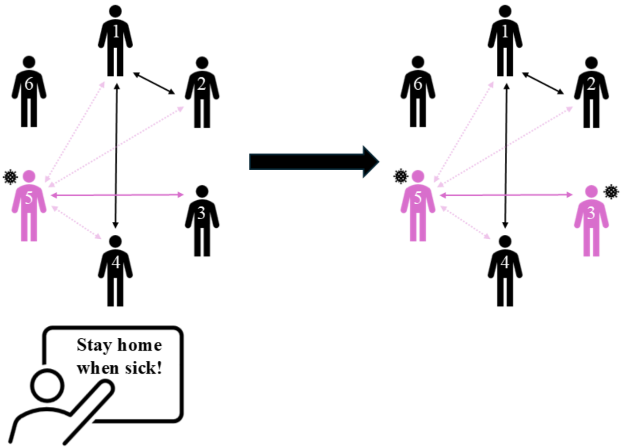
Example I: Intervention Affects Network and Transmission



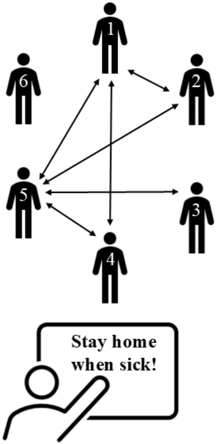
Example I: Intervention Affects Network and Transmission



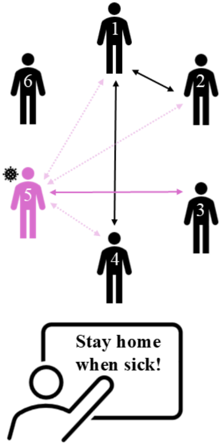
Example I: Intervention Affects Network and Transmission



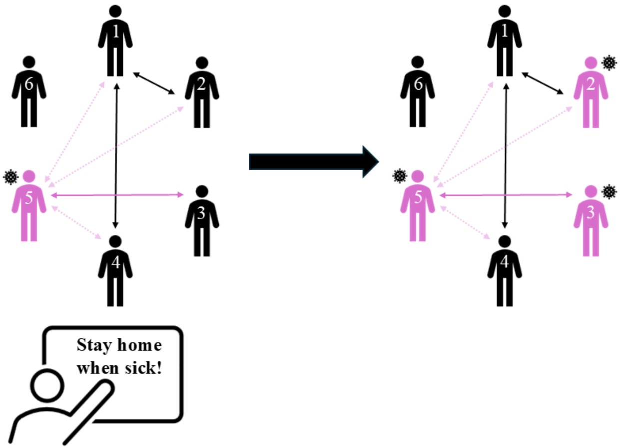
Example II: Intervention Affects Network, Not Transmission



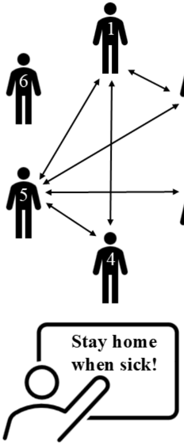
Example II: Intervention Affects Network, Not Transmission



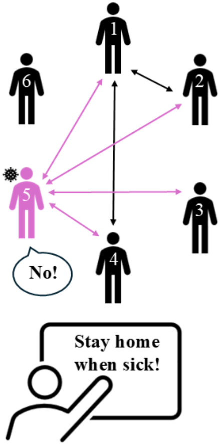
Example II: Intervention Affects Network, Not Transmission



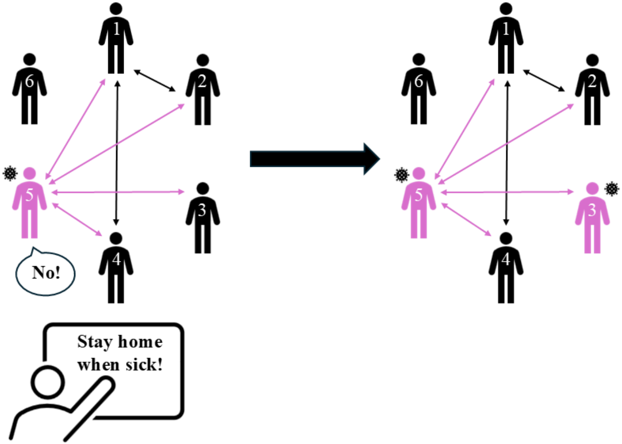
Example III: Intervention Affects Transmission, Not Network



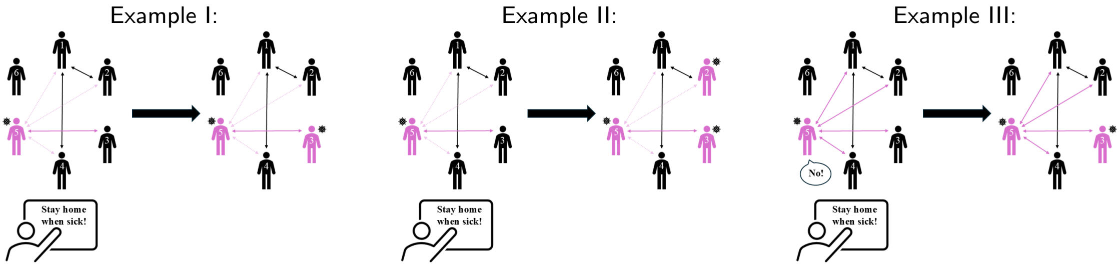
Example III: Intervention Affects Transmission, Not Network



Example III: Intervention Affects Transmission, Not Network



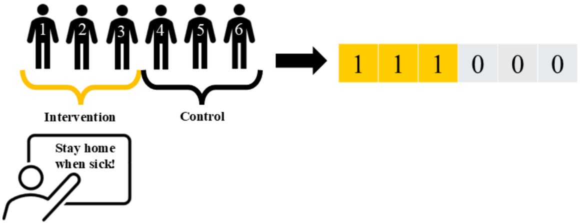
Central Question: Does the Intervention Affect Transmission of ILI?



In Examples I and III, the answer is yes

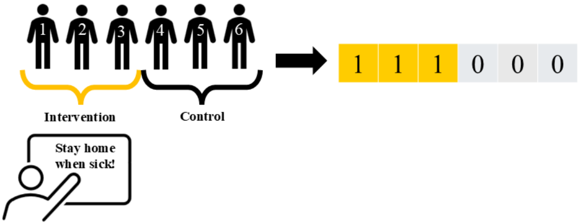
eX-FLU Observed Data

- Baseline randomization assignments
  $\mathbf{Z} = (Z_1, \ldots, Z_n) \in \mathcal{Z}$, for
  $Z_i = \mathbb{1}(\text{student } i \text{ gets intervention})$



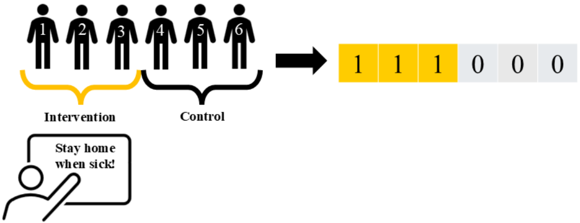
eX-FLU Observed Data

- Baseline randomization assignments $\mathbf{Z} = (Z_1, \ldots, Z_n) \in \mathcal{Z}$, for $Z_i = \mathbb{1}(\text{student } i \text{ gets intervention})$
  - randomization distribution $r(\mathbf{z})$



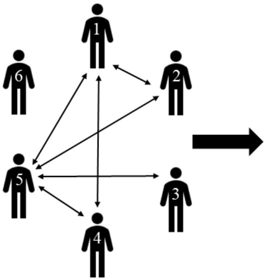
eX-FLU Observed Data

- Baseline randomization assignments $\mathbf{Z} = (Z_1, \ldots, Z_n) \in \mathcal{Z}$, for $Z_i = \mathbb{1}(\text{student } i \text{ gets intervention})$
  - randomization distribution $r(\mathbf{z})$
- At weeks $k = 1, \ldots, \tau$, we observe:



eX-FLU Observed Data

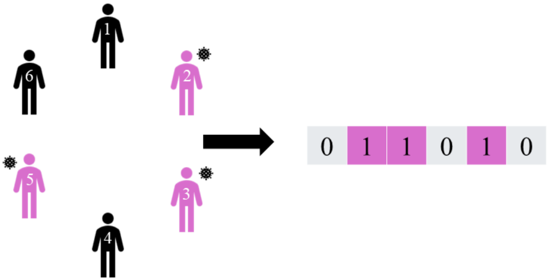
- Baseline randomization assignments $\mathbf{Z} = (Z_1, \ldots, Z_n) \in \mathcal{Z}$, for $Z_i = \mathbb{1}(\text{student } i \text{ gets intervention})$
  - randomization distribution $r(\mathbf{z})$
- At weeks $k = 1, \ldots, \tau$, we observe:
  - networks $\mathbf{A}^k = [A_{ij}^k]$, for $A_{ij}^k = \mathbb{1}(\text{students } i, j \text{ in contact at week } k)$



	1	2	3	4	5	6
1	0	1	0	1	1	0
2	1	0	0	0	1	0
3	0	0	0	0	1	0
4	1	0	0	0	1	0
5	1	1	1	1	0	0
6	0	0	0	0	0	0

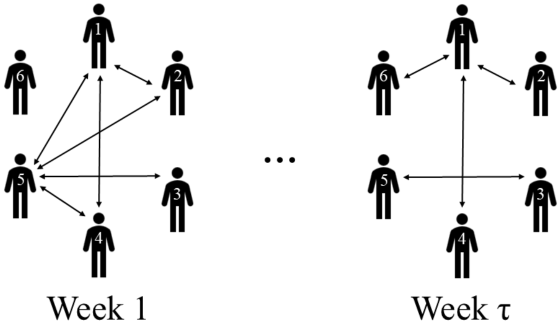
eX-FLU Observed Data

- Baseline randomization assignments $\mathbf{Z} = (Z_1, \ldots, Z_n) \in \mathcal{Z}$, for $Z_i = \mathbb{1}(\text{student } i \text{ gets intervention})$
  - randomization distribution $r(\mathbf{z})$
- At weeks $k = 1, \ldots, \tau$, we observe:
  - networks $\mathbf{A}^k = [A_{ij}^k]$, for $A_{ij}^k = \mathbb{1}(\text{students } i, j \text{ in contact at week } k)$
  - infections $\mathbf{Y}^k = (Y_1^k, \ldots, Y_n^k)$, for $Y_i^k = \mathbb{1}(\text{student } i \text{ infected at week } k)$



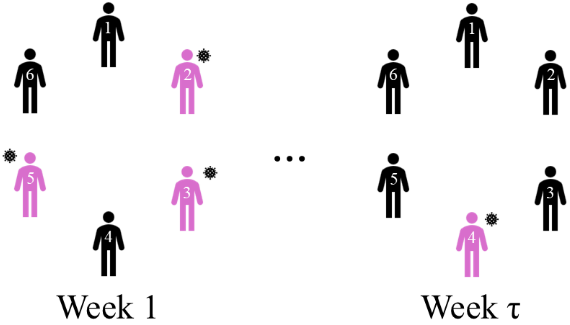
eX-FLU Observed Data

- Baseline randomization assignments $\mathbf{Z} = (Z_1, \ldots, Z_n) \in \mathcal{Z}$, for $Z_i = \mathbb{1}(\text{student } i \text{ gets intervention})$
  - randomization distribution $r(\mathbf{z})$
- At weeks $k = 1, \ldots, \tau$, we observe:
  - networks $\mathbf{A}^k = [A_{ij}^k]$, for $A_{ij}^k = \mathbb{1}(\text{students } i, j \text{ in contact at week } k)$
  - infections $\mathbf{Y}^k = (Y_1^k, \ldots, Y_n^k)$, for $Y_i^k = \mathbb{1}(\text{student } i \text{ infected at week } k)$
- network history: $\overline{\mathbf{A}} = \{\mathbf{A}^k\}_{k=1}^{\tau}$



eX-FLU Observed Data

- Baseline randomization assignments $\mathbf{Z} = (Z_1, \ldots, Z_n) \in \mathcal{Z}$, for $Z_i = \mathbb{1}(\text{student } i \text{ gets intervention})$
  - randomization distribution $r(\mathbf{z})$
- At weeks $k = 1, \ldots, \tau$, we observe:
  - networks $\mathbf{A}^k = [A^k_{ij}]$, for $A^k_{ij} = \mathbb{1}(\text{students } i, j \text{ in contact at week } k)$
  - infections $\mathbf{Y}^k = (Y^k_1, \ldots, Y^k_n)$, for $Y^k_i = \mathbb{1}(\text{student } i \text{ infected at week } k)$
- network history: $\overline{\mathbf{A}} = \{\mathbf{A}^k\}_{k=1}^{\tau}$
- infection history: $\overline{\mathbf{Y}} = \{\mathbf{Y}^k\}_{k=1}^{\tau}$



eX-FLU Potential Outcomes

Potential outcomes are defined for the networks and ILI infections:

$$\overline{\mathbf{A}}(\mathbf{z}), \quad \overline{\mathbf{Y}}(\mathbf{z}) \quad \text{for } \mathbf{z} \in \mathcal{Z}$$

- $\overline{\mathbf{A}}(\mathbf{z})$: sequence of networks that **would** occur if students **were** encouraged to isolate according to $\mathbf{z}$
- $\overline{\mathbf{Y}}(\mathbf{z})$: sequence of infection statuses that **would** occur if students **were** encouraged to isolate according to $\mathbf{z}$

eX-FLU Potential Outcomes

Potential outcomes are defined for the networks and ILI infections:

$$\overline{\mathbf{A}}(\mathbf{z}), \quad \overline{\mathbf{Y}}(\mathbf{z}) \quad \text{for } \mathbf{z} \in \mathcal{Z}$$

- $\overline{\mathbf{A}}(\mathbf{z})$: sequence of networks that **would** occur if students **were** encouraged to isolate according to $\mathbf{z}$
- $\overline{\mathbf{Y}}(\mathbf{z})$: sequence of infection statuses that **would** occur if students **were** encouraged to isolate according to $\mathbf{z}$
- Both of these depend on the entire vector $\mathbf{z}$
  - potential outcomes for one student depend on intervention assignment of other students (**interference**) (Hudgens and Halloran, 2008)

eX-FLU Potential Outcomes

Potential outcomes are defined for the networks and ILI infections:

$$\overline{\mathbf{A}}(\mathbf{z}), \quad \overline{\mathbf{Y}}(\mathbf{z}) \quad \text{for } \mathbf{z} \in \mathcal{Z}$$

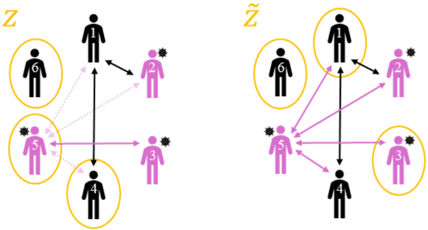
- $\overline{\mathbf{A}}(\mathbf{z})$: sequence of networks that **would** occur if students **were** encouraged to isolate according to $\mathbf{z}$
- $\overline{\mathbf{Y}}(\mathbf{z})$: sequence of infection statuses that **would** occur if students **were** encouraged to isolate according to $\mathbf{z}$
- Both of these depend on the entire vector $\mathbf{z}$
  - potential outcomes for one student depend on intervention assignment of other students (**interference**) (Hudgens and Halloran, 2008)
- Assume **causal consistency**:

$$\overline{\mathbf{A}} = \overline{\mathbf{A}}(\mathbf{Z}) \text{ and } \overline{\mathbf{Y}} = \overline{\mathbf{Y}}(\mathbf{Z})$$

eX-FLU Null Hypotheses

$$H_0^Y : \overline{\mathbf{Y}}(\mathbf{z}) = \overline{\mathbf{Y}}(\widetilde{\mathbf{z}}) \text{ for all } \mathbf{z}, \widetilde{\mathbf{z}} \in \mathcal{Z}$$

("the intervention has no effect on the infections")



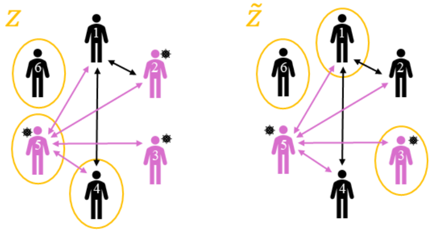
eX-FLU Null Hypotheses

$$H_0^Y : \overline{\mathbf{Y}}(\mathbf{z}) = \overline{\mathbf{Y}}(\widetilde{\mathbf{z}}) \text{ for all } \mathbf{z}, \widetilde{\mathbf{z}} \in \mathcal{Z}$$

("the intervention has no effect on the infections")

$$H_0^A : \overline{\mathbf{A}}(\mathbf{z}) = \overline{\mathbf{A}}(\widetilde{\mathbf{z}}) \text{ for all } \mathbf{z}, \widetilde{\mathbf{z}} \in \mathcal{Z}$$

("the intervention has no effect the networks")



eX-FLU Null Hypotheses

$$H_0^Y : \overline{\mathbf{Y}}(\mathbf{z}) = \overline{\mathbf{Y}}(\widetilde{\mathbf{z}}) \text{ for all } \mathbf{z}, \widetilde{\mathbf{z}} \in \mathcal{Z}$$

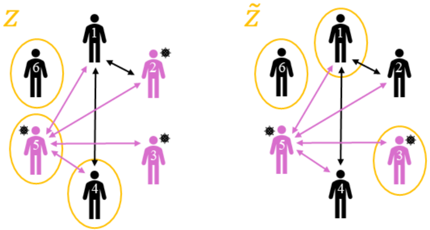
("the intervention has no effect on the infections")

$$H_0^A : \overline{\mathbf{A}}(\mathbf{z}) = \overline{\mathbf{A}}(\widetilde{\mathbf{z}}) \text{ for all } \mathbf{z}, \widetilde{\mathbf{z}} \in \mathcal{Z}$$

("the intervention has no effect the networks")

$$H_0^\sharp = H_0^Y \cap H_0^A : \overline{\mathbf{Y}}(\mathbf{z}) = \overline{\mathbf{Y}}(\widetilde{\mathbf{z}}) \text{ and } \overline{\mathbf{A}}(\mathbf{z}) = \overline{\mathbf{A}}(\widetilde{\mathbf{z}}) \text{ for all } \mathbf{z}, \widetilde{\mathbf{z}} \in \mathcal{Z}$$

("the intervention has no effect on the infections or on the networks")

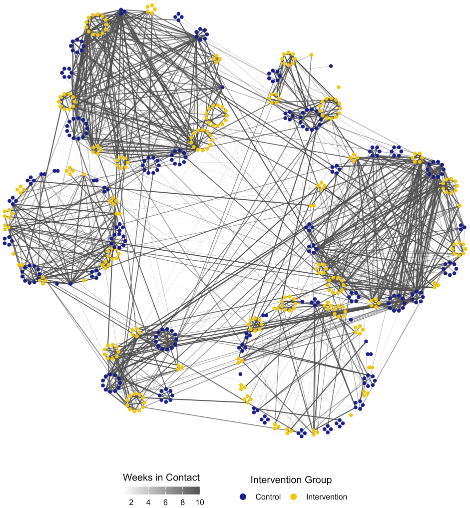


eX-FLU Null Hypotheses

$$H_0^Y : \overline{\mathbf{Y}}(\mathbf{z}) = \overline{\mathbf{Y}}(\widetilde{\mathbf{z}}) \text{ for all } \mathbf{z}, \widetilde{\mathbf{z}} \in \mathcal{Z}$$
$$\text{(“the intervention has no effect on the infections”)}$$
$$H_0^A : \overline{\mathbf{A}}(\mathbf{z}) = \overline{\mathbf{A}}(\widetilde{\mathbf{z}}) \text{ for all } \mathbf{z}, \widetilde{\mathbf{z}} \in \mathcal{Z}$$
$$\text{(“the intervention has no effect the networks”)}$$
$$H_0^\sharp = H_0^Y \cap H_0^A : \overline{\mathbf{Y}}(\mathbf{z}) = \overline{\mathbf{Y}}(\widetilde{\mathbf{z}}) \text{ and } \overline{\mathbf{A}}(\mathbf{z}) = \overline{\mathbf{A}}(\widetilde{\mathbf{z}}) \text{ for all } \mathbf{z}, \widetilde{\mathbf{z}} \in \mathcal{Z}$$
$$\text{(“the intervention has no effect on the infections or on the networks”)}$$

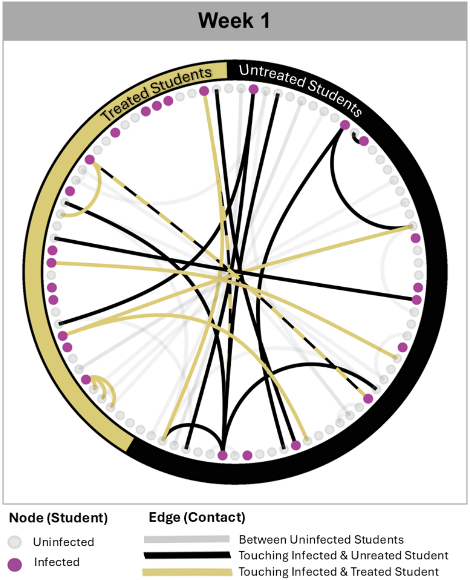
Previous analyses of the eX-FLU trial (Alexandria et al., 2023) tested $H_0^\sharp$, but no analyses have tested H_0^Y

eX-FLU: Descriptive Results



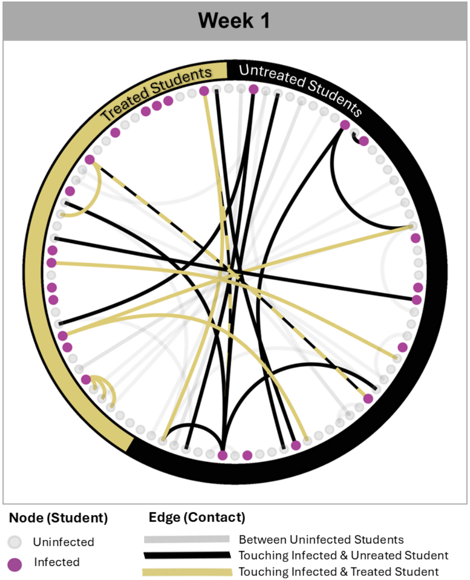
eX-FLU: Descriptive Results

- 93 (out of 579) students with at least one infection



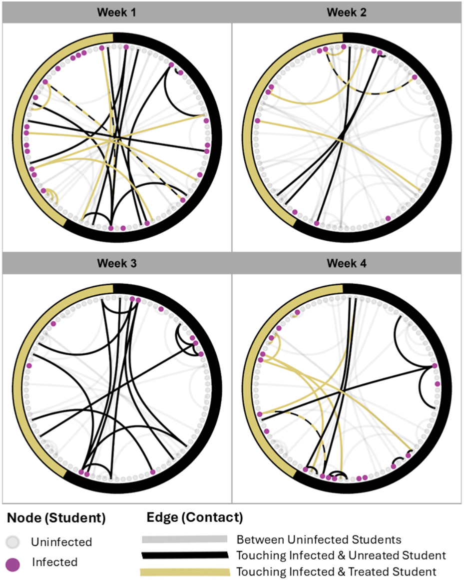
eX-FLU: Descriptive Results

- 93 (out of 579) students with at least one infection
- Bold edges represent possible transmission events



eX-FLU: Descriptive Results

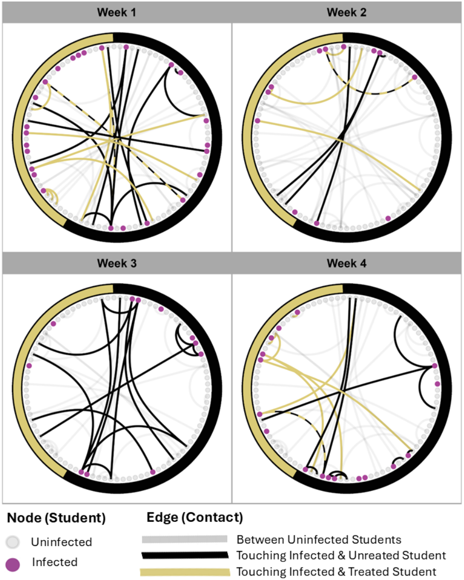
- 93 (out of 579) students with at least one infection
- Bold edges represent possible transmission events



eX-FLU: Descriptive Results

- 93 (out of 579) students with at least one infection
- Bold edges represent possible transmission events
  - Proportion of possible transmission events attributable to students in the intervention group:

$$T = \frac{\text{number of yellow edges}}{\text{total number of edges}} = \frac{\sum_{k=2}^{\tau}\sum_{i=1}^{n}\sum_{j>i} Z_i Y_i^{k-1} A_{ij}^{k-1}}{\sum_{k=2}^{\tau}\sum_{i=1}^{n}\sum_{j>i} Y_i^{k-1} A_{ij}^{k-1}} = 0.359$$



Testing $H_0^\sharp$

"How unlikely are the observed data under $H_0^\sharp$?"

Testing $H_0^\sharp$

"How unlikely are the observed data under $H_0^\sharp$?" We can answer this using a **randomization test** (Fisher, 1936; Alexandria et al., 2023; Zhang and Zhao, 2023; Ritzwoller et al., 2024)

Testing $H_0^\sharp$

"How unlikely are the observed data under $H_0^\sharp$?" We can answer this using a **randomization test** (Fisher, 1936; Alexandria et al., 2023; Zhang and Zhao, 2023; Ritzwoller et al., 2024)

- Choose a test statistic $T = T(\mathbf{Z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$

Testing $H_0^\sharp$

"How unlikely are the observed data under $H_0^\sharp$?" We can answer this using a **randomization test** (Fisher, 1936; Alexandria et al., 2023; Zhang and Zhao, 2023; Ritzwoller et al., 2024)

- Choose a test statistic $T = T(\mathbf{Z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
- Find the distribution of T under $H_0^\sharp$

Testing $H_0^\sharp$

"How unlikely are the observed data under $H_0^\sharp$?" We can answer this using a **randomization test** (Fisher, 1936; Alexandria et al., 2023; Zhang and Zhao, 2023; Ritzwoller et al., 2024)

- Choose a test statistic $T = T(\mathbf{Z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
- Find the distribution of T under $H_0^\sharp$

Randomization	Networks	Infections	Test Statistic
$\mathbf{z}_1$			
$\mathbf{z}_2$			
$\dots$			
$\mathbf{z}_{\lvert\mathcal{Z}\rvert}$			

Testing $H_0^\sharp$

"How unlikely are the observed data under $H_0^\sharp$?" We can answer this using a **randomization test** (Fisher, 1936; Alexandria et al., 2023; Zhang and Zhao, 2023; Ritzwoller et al., 2024)

- Choose a test statistic $T = T(\mathbf{Z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
- Find the distribution of T under $H_0^\sharp$

Randomization	**Networks**	**Infections**	**Test Statistic**
$\mathbf{z}_1$	$\overline{\mathbf{A}}(\mathbf{z}_1) =?$	$\overline{\mathbf{Y}}(\mathbf{z}_1) =?$	$T(\mathbf{z}_1, ?, ?)$
$\mathbf{z}_2$	$\overline{\mathbf{A}}(\mathbf{z}_2) =?$	$\overline{\mathbf{Y}}(\mathbf{z}_1) =?$	$T(\mathbf{z}_1, ?, ?)$
...	...	...	...
$\mathbf{z}_{\lvert\mathcal{Z}\rvert}$	$\overline{\mathbf{A}}(\mathbf{z}_{\lvert\mathcal{Z}\rvert}) =?$	$\overline{\mathbf{Y}}(\mathbf{z}_{\lvert\mathcal{Z}\rvert}) =?$	$T(\mathbf{z}_{\lvert\mathcal{Z}\rvert}, ?, ?)$

Testing $H_0^\sharp$

"How unlikely are the observed data under $H_0^\sharp$?" We can answer this using a **randomization test** (Fisher, 1936; Alexandria et al., 2023; Zhang and Zhao, 2023; Ritzwoller et al., 2024)

- Choose a test statistic $T = T(\mathbf{Z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
- Find the distribution of T under $H_0^\sharp$

Randomization	Networks	Infections	Test Statistic
$\mathbf{z}_1$	$\overline{\mathbf{A}}(\mathbf{z}_1) = \overline{\mathbf{A}}$	$\overline{\mathbf{Y}}(\mathbf{z}_1) = \overline{\mathbf{Y}}$	$T(\mathbf{z}_1, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
$\mathbf{z}_2$	$\overline{\mathbf{A}}(\mathbf{z}_2) = \overline{\mathbf{A}}$	$\overline{\mathbf{Y}}(\mathbf{z}_2) = \overline{\mathbf{Y}}$	$T(\mathbf{z}_2, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
$\ldots$	$\ldots$	$\ldots$	$\ldots$
$\mathbf{z}_{\lvert\mathcal{Z}\rvert}$	$\overline{\mathbf{A}}(\mathbf{z}_{\lvert\mathcal{Z}\rvert}) = \overline{\mathbf{A}}$	$\overline{\mathbf{Y}}(\mathbf{z}_{\lvert\mathcal{Z}\rvert}) = \overline{\mathbf{Y}}$	$T(\mathbf{z}_{\lvert\mathcal{Z}\rvert}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$

Testing $H_0^\sharp$

"How unlikely are the observed data under $H_0^\sharp$?" We can answer this using a **randomization test** (Fisher, 1936; Alexandria et al., 2023; Zhang and Zhao, 2023; Ritzwoller et al., 2024)

- Choose a test statistic $T = T(\mathbf{Z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
- Find the distribution of T under $H_0^\sharp$
  - This is possible since $\overline{\mathbf{A}}(\mathbf{z}), \overline{\mathbf{Y}}(\mathbf{z})$ are **imputable** under $H_0^\sharp$

Randomization	**Networks**	**Infections**	**Test Statistic**
$\mathbf{z}_1$	$\overline{\mathbf{A}}(\mathbf{z}_1) = \overline{\mathbf{A}}$	$\overline{\mathbf{Y}}(\mathbf{z}_1) = \overline{\mathbf{Y}}$	$T(\mathbf{z}_1, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
$\mathbf{z}_2$	$\overline{\mathbf{A}}(\mathbf{z}_2) = \overline{\mathbf{A}}$	$\overline{\mathbf{Y}}(\mathbf{z}_2) = \overline{\mathbf{Y}}$	$T(\mathbf{z}_2, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
$\dots$	$\dots$	$\dots$	$\dots$
$\mathbf{z}_{\lvert\mathcal{Z}\rvert}$	$\overline{\mathbf{A}}(\mathbf{z}_{\lvert\mathcal{Z}\rvert}) = \overline{\mathbf{A}}$	$\overline{\mathbf{Y}}(\mathbf{z}_{\lvert\mathcal{Z}\rvert}) = \overline{\mathbf{Y}}$	$T(\mathbf{z}_{\lvert\mathcal{Z}\rvert}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$

Testing $H_0^\sharp$

"How unlikely are the observed data under $H_0^\sharp$?" We can answer this using a **randomization test** (Fisher, 1936; Alexandria et al., 2023; Zhang and Zhao, 2023; Ritzwoller et al., 2024)

- Choose a test statistic $T = T(\mathbf{Z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
- Find the distribution of T under $H_0^\sharp$
  - This is possible since $\overline{\mathbf{A}}(\mathbf{z}), \overline{\mathbf{Y}}(\mathbf{z})$ are **imputable** under $H_0^\sharp \implies H_0^\sharp$ is **sharp**

Randomization	Networks	Infections	Test Statistic
$\mathbf{z}_1$	$\overline{\mathbf{A}}(\mathbf{z}_1) = \overline{\mathbf{A}}$	$\overline{\mathbf{Y}}(\mathbf{z}_1) = \overline{\mathbf{Y}}$	$T(\mathbf{z}_1, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
$\mathbf{z}_2$	$\overline{\mathbf{A}}(\mathbf{z}_2) = \overline{\mathbf{A}}$	$\overline{\mathbf{Y}}(\mathbf{z}_2) = \overline{\mathbf{Y}}$	$T(\mathbf{z}_2, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
$\dots$	$\dots$	$\dots$	$\dots$
$\mathbf{z}_{\lvert\mathcal{Z}\rvert}$	$\overline{\mathbf{A}}(\mathbf{z}_{\lvert\mathcal{Z}\rvert}) = \overline{\mathbf{A}}$	$\overline{\mathbf{Y}}(\mathbf{z}_{\lvert\mathcal{Z}\rvert}) = \overline{\mathbf{Y}}$	$T(\mathbf{z}_{\lvert\mathcal{Z}\rvert}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$

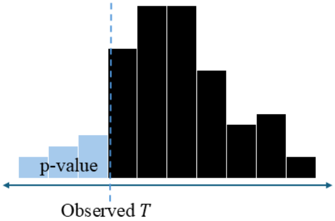
Testing $H_0^\sharp$

"How unlikely are the observed data under $H_0^\sharp$?" We can answer this using a **randomization test** (Fisher, 1936; Alexandria et al., 2023; Zhang and Zhao, 2023; Ritzwoller et al., 2024)

- Choose a test statistic $T = T(\mathbf{Z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
- Find the distribution of T under $H_0^\sharp$
  - This is possible since $\overline{\mathbf{A}}(\mathbf{z}), \overline{\mathbf{Y}}(\mathbf{z})$ are **imputable** under $H_0^\sharp \implies H_0^\sharp$ is **sharp**
- Compute a p-value: $\rho^\sharp = F_T(T|H_0^\sharp)$, where

$$F_T(t|H_0^\sharp) = \Pr\left(T \leq t|H_0^\sharp\right) = \sum_{\mathbf{z}\in\mathcal{Z}} r(\mathbf{z})\mathbb{1}\{T(\mathbf{z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}}) \leq t\}$$

  is the **cumulative distribution function** (CDF) of T



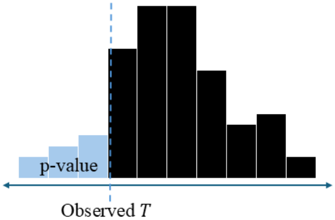
Testing $H_0^\sharp$

"How unlikely are the observed data under $H_0^\sharp$?" We can answer this using a **randomization test** (Fisher, 1936; Alexandria et al., 2023; Zhang and Zhao, 2023; Ritzwoller et al., 2024)

- Choose a test statistic $T = T(\mathbf{Z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
- Find the distribution of T under $H_0^\sharp$
  - This is possible since $\overline{\mathbf{A}}(\mathbf{z}), \overline{\mathbf{Y}}(\mathbf{z})$ are **imputable** under $H_0^\sharp \implies H_0^\sharp$ is **sharp**
- Compute a p-value: $\rho^\sharp = F_T(T|H_0^\sharp)$, where

$$F_T(t|H_0^\sharp) = \Pr\left(T \leq t|H_0^\sharp\right) = \sum_{\mathbf{z}\in\mathcal{Z}} r(\mathbf{z})\mathbb{1}\{T(\mathbf{z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}}) \leq t\}$$

  is the **cumulative distribution function** (CDF) of T
- Reject $H_0^\sharp$ if $\rho^\sharp \leq 0.05$



Testing $H_0^\sharp$

- This test of $H_0^\sharp$ controls the **type I error rate** exactly:

$$\Pr\left(\rho^\sharp \leq \alpha | H_0^\sharp\right) \leq \alpha$$

for any $\alpha \in [0, 1]$, for any sample size, and for any choice of test statistic

Testing $H_0^\sharp$

- This test of $H_0^\sharp$ controls the **type I error rate** exactly:

$$\Pr\left(\rho^\sharp \leq \alpha | H_0^\sharp\right) \leq \alpha$$

for any $\alpha \in [0, 1]$, for any sample size, and for any choice of test statistic

- The **power** of this test:

$$\Pr\left(\rho^\sharp \leq \alpha | H_1^\sharp\right)$$

depends on the choice of test statistic

Testing H_0^Y

"How unlikely are the observed data under H_0^Y?"

Testing H_0^Y

"How unlikely are the observed data under H_0^Y?"

- Choose a test statistic $T = T(\mathbf{Z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$

Testing H_0^Y

"How unlikely are the observed data under H_0^Y?"

- Choose a test statistic $T = T(\mathbf{Z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
- Find the distribution of T under H_0^Y

Testing H_0^Y

"How unlikely are the observed data under H_0^Y?"

- Choose a test statistic $T = T(\mathbf{Z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
- Find the distribution of T under H_0^Y

Randomization	**Networks**	**Infections**	**Test Statistic**
$\mathbf{z}_1$			
$\mathbf{z}_2$			
$\cdots$			
$\mathbf{z}_{\lvert\mathcal{Z}\rvert}$			

Testing H_0^Y

"How unlikely are the observed data under H_0^Y?"

- Choose a test statistic $T = T(\mathbf{Z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
- Find the distribution of T under H_0^Y

Randomization	Networks	Infections	Test Statistic
$\mathbf{z}_1$	$\overline{\mathbf{A}}(\mathbf{z}_1) =?$	$\overline{\mathbf{Y}}(\mathbf{z}_1) =?$	$T(\mathbf{z}_1, ?, ?)$
$\mathbf{z}_2$	$\overline{\mathbf{A}}(\mathbf{z}_2) =?$	$\overline{\mathbf{Y}}(\mathbf{z}_1) =?$	$T(\mathbf{z}_1, ?, ?)$
$\dots$	$\dots$	$\dots$	$\dots$
$\mathbf{z}_{\lvert\mathcal{Z}\rvert}$	$\overline{\mathbf{A}}(\mathbf{z}_{\lvert\mathcal{Z}\rvert}) =?$	$\overline{\mathbf{Y}}(\mathbf{z}_{\lvert\mathcal{Z}\rvert}) =?$	$T(\mathbf{z}_{\lvert\mathcal{Z}\rvert}, ?, ?)$

Testing H_0^Y

"How unlikely are the observed data under H_0^Y?"

- Choose a test statistic $T = T(\mathbf{Z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
- Find the distribution of T under H_0^Y

Randomization	Networks	Infections	Test Statistic
$\mathbf{z}_1$	$\overline{\mathbf{A}}(\mathbf{z}_1) = ?$	$\overline{\mathbf{Y}}(\mathbf{z}_1) = \overline{\mathbf{Y}}$	$T(\mathbf{z}_1, ?, \overline{\mathbf{Y}})$
$\mathbf{z}_2$	$\overline{\mathbf{A}}(\mathbf{z}_2) = ?$	$\overline{\mathbf{Y}}(\mathbf{z}_2) = \overline{\mathbf{Y}}$	$T(\mathbf{z}_2, ?, \overline{\mathbf{Y}})$
$\ldots$	$\ldots$	$\ldots$	$\ldots$
$\mathbf{z}_{\|\mathcal{Z}\|}$	$\overline{\mathbf{A}}(\mathbf{z}_{\|\mathcal{Z}\|}) = ?$	$\overline{\mathbf{Y}}(\mathbf{z}_{\|\mathcal{Z}\|}) = \overline{\mathbf{Y}}$	$T(\mathbf{z}_{\|\mathcal{Z}\|}, ?, \overline{\mathbf{Y}})$

Testing H_0^Y

"How unlikely are the observed data under H_0^Y?"

- Choose a test statistic $T = T(\mathbf{Z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
- Find the distribution of T under H_0^Y
  - This is not possible since $\overline{\mathbf{A}}(\mathbf{z})$ are **not imputable** under H_0^Y

Randomization	Networks	Infections	Test Statistic
$\mathbf{z}_1$	$\overline{\mathbf{A}}(\mathbf{z}_1) = ?$	$\overline{\mathbf{Y}}(\mathbf{z}_1) = \overline{\mathbf{Y}}$	$T(\mathbf{z}_1, ?, \overline{\mathbf{Y}})$
$\mathbf{z}_2$	$\overline{\mathbf{A}}(\mathbf{z}_2) = ?$	$\overline{\mathbf{Y}}(\mathbf{z}_2) = \overline{\mathbf{Y}}$	$T(\mathbf{z}_2, ?, \overline{\mathbf{Y}})$
$\ldots$	$\ldots$	$\ldots$	$\ldots$
$\mathbf{z}_{\lvert\mathcal{Z}\rvert}$	$\overline{\mathbf{A}}(\mathbf{z}_{\lvert\mathcal{Z}\rvert}) = ?$	$\overline{\mathbf{Y}}(\mathbf{z}_{\lvert\mathcal{Z}\rvert}) = \overline{\mathbf{Y}}$	$T(\mathbf{z}_{\lvert\mathcal{Z}\rvert}, ?, \overline{\mathbf{Y}})$

Testing H_0^Y

"How unlikely are the observed data under H_0^Y?"

- Choose a test statistic $T = T(\mathbf{Z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}})$
- Find the distribution of T under H_0^Y
  - This is not possible since $\overline{\mathbf{A}}(\mathbf{z})$ are **not imputable** under $H_0^Y \implies H_0^Y$ is **nonsharp**

Randomization	Networks	Infections	Test Statistic
$\mathbf{z}_1$	$\overline{\mathbf{A}}(\mathbf{z}_1) = ?$	$\overline{\mathbf{Y}}(\mathbf{z}_1) = \overline{\mathbf{Y}}$	$T(\mathbf{z}_1, ?, \overline{\mathbf{Y}})$
$\mathbf{z}_2$	$\overline{\mathbf{A}}(\mathbf{z}_2) = ?$	$\overline{\mathbf{Y}}(\mathbf{z}_2) = \overline{\mathbf{Y}}$	$T(\mathbf{z}_2, ?, \overline{\mathbf{Y}})$
$\dots$	$\dots$	$\dots$	$\dots$
$\mathbf{z}_{\lvert\mathcal{Z}\rvert}$	$\overline{\mathbf{A}}(\mathbf{z}_{\lvert\mathcal{Z}\rvert}) = ?$	$\overline{\mathbf{Y}}(\mathbf{z}_{\lvert\mathcal{Z}\rvert}) = \overline{\mathbf{Y}}$	$T(\mathbf{z}_{\lvert\mathcal{Z}\rvert}, ?, \overline{\mathbf{Y}})$

Testing H_0^Y

- **Problem**: H_0^Y is nonsharp

Testing H_0^Y

- **Problem**: H_0^Y is nonsharp
  $\implies$ we don't know $F_T(t|H_0^Y)$ and thus cannot compute a p-value testing H_0^Y

Testing H_0^Y

- **Problem**: H_0^Y is nonsharp
  $\Longrightarrow$ we don't know $F_T(t|H_0^Y)$ and thus cannot compute a p-value testing H_0^Y
- **Solution**: assume the potential networks are **stochastic**

Testing H_0^Y

- **Problem**: H_0^Y is nonsharp
  $\implies$ we don't know $F_T(t|H_0^Y)$ and thus cannot compute a p-value testing H_0^Y
- **Solution**: assume the potential networks are **stochastic**
  - Define $q(\overline{\mathbf{a}}, \mathbf{z}) = \Pr\{\overline{\mathbf{A}}(\mathbf{z}) = \overline{\mathbf{a}}\}$, the **probability mass function** (PMF) of $\overline{\mathbf{A}}(\mathbf{z})$

Testing H_0^Y

- **Problem**: H_0^Y is nonsharp
  $\implies$ we don't know $F_T(t|H_0^Y)$ and thus cannot compute a p-value testing H_0^Y
- **Solution**: assume the potential networks are **stochastic**
  - Define $q(\bar{\mathbf{a}}, \mathbf{z}) = \Pr\{\overline{\mathbf{A}}(\mathbf{z}) = \bar{\mathbf{a}}\}$, the **probability mass function** (PMF) of $\overline{\mathbf{A}}(\mathbf{z})$
  - Then

$$F_T(t|H_0^Y) = \Pr\left(T \leq t|H_0^Y\right) = \sum_{\mathbf{z}\in\mathcal{Z}} \sum_{\bar{\mathbf{a}}\in\mathcal{A}} r(\mathbf{z})q(\bar{\mathbf{a}}, \mathbf{z})\mathbb{1}\{T(\mathbf{z}, \bar{\mathbf{a}}, \overline{\mathbf{Y}}) \leq t\}$$

Testing H_0^Y

- **Problem**: H_0^Y is nonsharp
  $\implies$ we don't know $F_T(t|H_0^Y)$ and thus cannot compute a p-value testing H_0^Y
- **Solution**: assume the potential networks are **stochastic**
  - Define $q(\bar{\mathbf{a}}, \mathbf{z}) = \Pr\{\overline{\mathbf{A}}(\mathbf{z}) = \bar{\mathbf{a}}\}$, the **probability mass function** (PMF) of $\overline{\mathbf{A}}(\mathbf{z})$
  - Then

$$F_T(t|H_0^Y) = \Pr\left(T \leq t|H_0^Y\right) = \sum_{\mathbf{z}\in\mathcal{Z}} \sum_{\bar{\mathbf{a}}\in\mathcal{A}} r(\mathbf{z}) q(\bar{\mathbf{a}}, \mathbf{z}) \mathbb{1}\{T(\mathbf{z}, \bar{\mathbf{a}}, \overline{\mathbf{Y}}) \leq t\}$$

  - p-value: $\rho^Y = F_T(T|H_0^Y)$

Testing H_0^Y

- **Problem**: H_0^Y is nonsharp

  $\implies$ we don't know $F_T(t|H_0^Y)$ and thus cannot compute a p-value testing H_0^Y
- **Solution**: assume the potential networks are **stochastic**
  - Define $q(\overline{\mathbf{a}}, \mathbf{z}) = \Pr\{\overline{\mathbf{A}}(\mathbf{z}) = \overline{\mathbf{a}}\}$, the **probability mass function** (PMF) of $\overline{\mathbf{A}}(\mathbf{z})$
  - Then

$$F_T(t|H_0^Y) = \Pr\left(T \leq t|H_0^Y\right) = \sum_{\mathbf{z}\in\mathcal{Z}} \sum_{\overline{\mathbf{a}}\in\mathcal{A}} r(\mathbf{z})q(\overline{\mathbf{a}}, \mathbf{z})\mathbb{1}\{T(\mathbf{z}, \overline{\mathbf{a}}, \overline{\mathbf{Y}}) \leq t\}$$

  - p-value: $\rho^Y = F_T(T|H_0^Y)$
  - If/when q is unknown, estimate it

Testing H_0^Y

- **Problem**: H_0^Y is nonsharp
  $\implies$ we don't know $F_T(t|H_0^Y)$ and thus cannot compute a p-value testing H_0^Y
- **Solution**: assume the potential networks are **stochastic**
  - Define $q(\bar{\mathbf{a}}, \mathbf{z}) = \Pr\{\overline{\mathbf{A}}(\mathbf{z}) = \bar{\mathbf{a}}\}$, the **probability mass function** (PMF) of $\overline{\mathbf{A}}(\mathbf{z})$
  - Then

$$F_T(t|H_0^Y) = \Pr\left(T \leq t|H_0^Y\right) = \sum_{\mathbf{z}\in\mathcal{Z}} \sum_{\bar{\mathbf{a}}\in\mathcal{A}} r(\mathbf{z})q(\bar{\mathbf{a}}, \mathbf{z})\mathbb{1}\{T(\mathbf{z}, \bar{\mathbf{a}}, \overline{\mathbf{Y}}) \leq t\}$$

  - p-value: $\rho^Y = F_T(T|H_0^Y)$
  - If/when q is unknown, estimate it
    - ★ E.g., using a **Separable Temporal Exponential Random Graph Model**

Testing H_0^Y

- **Problem**: H_0^Y is nonsharp
  $\implies$ we don't know $F_T(t|H_0^Y)$ and thus cannot compute a p-value testing H_0^Y
- **Solution**: assume the potential networks are **stochastic**
  - Define $q(\overline{\mathbf{a}}, \mathbf{z}) = \Pr\{\overline{\mathbf{A}}(\mathbf{z}) = \overline{\mathbf{a}}\}$, the **probability mass function** (PMF) of $\overline{\mathbf{A}}(\mathbf{z})$
  - Then

$$F_T(t|H_0^Y) = \Pr\left(T \leq t|H_0^Y\right) = \sum_{\mathbf{z}\in\mathcal{Z}} \sum_{\overline{\mathbf{a}}\in\mathcal{A}} r(\mathbf{z})q(\overline{\mathbf{a}}, \mathbf{z})\mathbb{1}\{T(\mathbf{z}, \overline{\mathbf{a}}, \overline{\mathbf{Y}}) \leq t\}$$

  - p-value: $\rho^Y = F_T(T|H_0^Y)$
  - If/when q is unknown, estimate it
    - ⋆ E.g., using a **Separable Temporal Exponential Random Graph Model**
  - If/when the double sum $\sum_{\mathbf{z}\in\mathcal{Z}} \sum_{\overline{\mathbf{a}}\in\mathcal{A}}$ cannot be computed exactly, approximate it using a Monte-Carlo method

Testing H_0^Y

Testing Procedure:

1. Estimate the unknown parameter θ in the PMF $q(\bar{\mathbf{a}}, \mathbf{z}; \theta)$ of $\overline{\mathbf{A}}(\mathbf{z})$

Testing H_0^Y

Testing Procedure:

1. Estimate the unknown parameter θ in the PMF $q(\bar{\mathbf{a}}, \mathbf{z}; \theta)$ of $\overline{\mathbf{A}}(\mathbf{z})$
2. Approximate $F_T(t|H_0^Y)$ using B Monte-Carlo replicates:
   for $b = 1$ to B **do:**

Testing H_0^Y

Testing Procedure:

1. Estimate the unknown parameter θ in the PMF $q(\bar{\mathbf{a}}, \mathbf{z}; \theta)$ of $\overline{\mathbf{A}}(\mathbf{z})$
2. Approximate $F_T(t|H_0^Y)$ using B Monte-Carlo replicates:
   for $b = 1$ to B **do:**
   1. generate randomization vector $\mathbf{z}_b$

Testing H_0^Y

Testing Procedure:

1. Estimate the unknown parameter θ in the PMF $q(\overline{\mathbf{a}}, \mathbf{z}; \theta)$ of $\overline{\mathbf{A}}(\mathbf{z})$
2. Approximate $F_T(t|H_0^Y)$ using B Monte-Carlo replicates:
   for $b = 1$ to B **do:**
   1. generate randomization vector $\mathbf{z}_b$
   2. generate networks $\overline{\mathbf{a}}_b$ from $q(\overline{\mathbf{a}}, \mathbf{z}_b; \widehat{\theta})$

Testing H_0^Y

Testing Procedure:

1. Estimate the unknown parameter θ in the PMF $q(\bar{\mathbf{a}}, \mathbf{z}; \theta)$ of $\overline{\mathbf{A}}(\mathbf{z})$
2. Approximate $F_T(t|H_0^Y)$ using B Monte-Carlo replicates:
   for $b = 1$ to B **do:**
   1. generate randomization vector $\mathbf{z}_b$
   2. generate networks $\bar{\mathbf{a}}_b$ from $q(\bar{\mathbf{a}}, \mathbf{z}_b; \widehat{\theta})$
   3. compute the test statistic $T_b = T(\mathbf{z}_b, \bar{\mathbf{a}}_b, \mathbf{Y})$

Testing H_0^Y

Testing Procedure:

1. Estimate the unknown parameter θ in the PMF $q(\bar{\mathbf{a}}, \mathbf{z}; \theta)$ of $\overline{\mathbf{A}}(\mathbf{z})$
2. Approximate $F_T(t|H_0^Y)$ using B Monte-Carlo replicates:
   for $b = 1$ to B **do:**
   1. generate randomization vector $\mathbf{z}_b$
   2. generate networks $\bar{\mathbf{a}}_b$ from $q(\bar{\mathbf{a}}, \mathbf{z}_b; \widehat{\theta})$
   3. compute the test statistic $T_b = T(\mathbf{z}_b, \bar{\mathbf{a}}_b, \mathbf{Y})$
3. compute the plug-in p-value testing H_0^Y:

$$\widehat{\rho}_B^Y = B^{-1} \sum_{b=1}^{B} \mathbb{1}(T_b \leq T)$$

Testing H_0^Y

Testing Procedure:

1. Estimate the unknown parameter θ in the PMF $q(\bar{\mathbf{a}}, \mathbf{z}; \theta)$ of $\overline{\mathbf{A}}(\mathbf{z})$
2. Approximate $F_T(t|H_0^Y)$ using B Monte-Carlo replicates:
   for $b = 1$ to B **do:**
   1. generate randomization vector $\mathbf{z}_b$
   2. generate networks $\bar{\mathbf{a}}_b$ from $q(\bar{\mathbf{a}}, \mathbf{z}_b; \widehat{\theta})$
   3. compute the test statistic $T_b = T(\mathbf{z}_b, \bar{\mathbf{a}}_b, \mathbf{Y})$
3. compute the plug-in p-value testing H_0^Y:

$$\widehat{\rho}_B^Y = B^{-1} \sum_{b=1}^{B} \mathbb{1}(T_b \leq T)$$

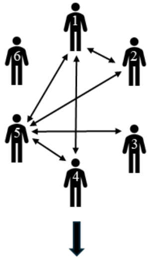
Proposition: This procedure will **asymptotically** control the type I error rate

Exponential Family Random Graph Models

Goal: model a network ($\mathbf{A}$) given covariates ($\mathbf{X}$)

Exponential Family Random Graph Models

Goal: model a network (**A**) given covariates (**X**)



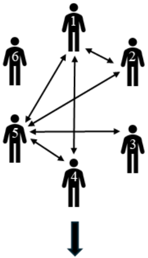
	1	2	3	4	5	6
1	0	1	0	1	1	0
2	1	0	0	0	1	0
3	0	0	0	0	1	0
4	1	0	0	0	1	0
5	1	1	1	1	0	0
6	0	0	0	0	0	0

Exponential Family Random Graph Models

Goal: model a network (**A**) given covariates (**X**)

An **ERGM** assumes:

$$\Pr_{\theta}(\mathbf{A} = \mathbf{a} | \mathbf{X} = \mathbf{x}) = \frac{\exp\{\theta \cdot \mathbf{g}(\mathbf{a}, \mathbf{x})\}}{\kappa(\theta, \mathcal{A}, \mathbf{x})}$$



	1	2	3	4	5	6
1	0	1	0	1	1	0
2	1	0	0	0	1	0
3	0	0	0	0	1	0
4	1	0	0	0	1	0
5	1	1	1	1	0	0
6	0	0	0	0	0	0

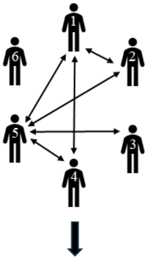
Exponential Family Random Graph Models

Goal: model a network ($\mathbf{A}$) given covariates ($\mathbf{X}$)

An **ERGM** assumes:

$$\Pr_{\theta}(\mathbf{A} = \mathbf{a} | \mathbf{X} = \mathbf{x}) = \frac{\exp\{\theta \cdot \mathbf{g}(\mathbf{a}, \mathbf{x})\}}{\kappa(\theta, \mathcal{A}, \mathbf{x})}$$

- normalizing constant: $\kappa(\theta, \mathcal{A}, \mathbf{x}) = \sum_{\mathbf{a} \in \mathcal{A}} \exp\{\theta \cdot \mathbf{g}(\mathbf{a}, \mathbf{x})\}$



	1	2	3	4	5	6
1	0	1	0	1	1	0
2	1	0	0	0	1	0
3	0	0	0	0	1	0
4	1	0	0	0	1	0
5	1	1	1	1	0	0
6	0	0	0	0	0	0

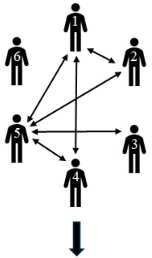
Exponential Family Random Graph Models

Goal: model a network ($\mathbf{A}$) given covariates ($\mathbf{X}$)

An **ERGM** assumes:

$$\Pr_{\theta}(\mathbf{A} = \mathbf{a} | \mathbf{X} = \mathbf{x}) = \frac{\exp\{\theta \cdot \mathbf{g}(\mathbf{a}, \mathbf{x})\}}{\kappa(\theta, \mathcal{A}, \mathbf{x})}$$

- normalizing constant: $\kappa(\theta, \mathcal{A}, \mathbf{x}) = \sum_{\mathbf{a} \in \mathcal{A}} \exp\{\theta \cdot \mathbf{g}(\mathbf{a}, \mathbf{x})\}$
- $\mathcal{A}$: space of possible networks



	1	2	3	4	5	6
1	0	1	0	1	1	0
2	1	0	0	0	1	0
3	0	0	0	0	1	0
4	1	0	0	0	1	0
5	1	1	1	1	0	0
6	0	0	0	0	0	0

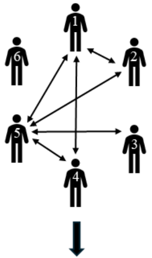
Exponential Family Random Graph Models

Goal: model a network ($\mathbf{A}$) given covariates ($\mathbf{X}$)

An **ERGM** assumes:

$$\Pr_{\theta}(\mathbf{A} = \mathbf{a} | \mathbf{X} = \mathbf{x}) = \frac{\exp\{\theta \cdot \mathbf{g}(\mathbf{a}, \mathbf{x})\}}{\kappa(\theta, \mathcal{A}, \mathbf{x})}$$

- normalizing constant:
  $\kappa(\theta, \mathcal{A}, \mathbf{x}) = \sum_{\mathbf{a} \in \mathcal{A}} \exp\{\theta \cdot \mathbf{g}(\mathbf{a}, \mathbf{x})\}$
- $\mathcal{A}$: space of possible networks
- θ: model coefficients



	1	2	3	4	5	6
1	0	1	0	1	1	0
2	1	0	0	0	1	0
3	0	0	0	0	1	0
4	1	0	0	0	1	0
5	1	1	1	1	0	0
6	0	0	0	0	0	0

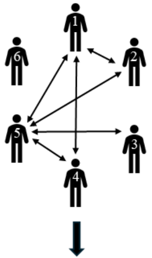
Exponential Family Random Graph Models

Goal: model a network ($\mathbf{A}$) given covariates ($\mathbf{X}$)

An **ERGM** assumes:

$$\Pr_{\theta}(\mathbf{A} = \mathbf{a} | \mathbf{X} = \mathbf{x}) = \frac{\exp\{\theta \cdot \mathbf{g}(\mathbf{a}, \mathbf{x})\}}{\kappa(\theta, \mathcal{A}, \mathbf{x})}$$

- normalizing constant: $\kappa(\theta, \mathcal{A}, \mathbf{x}) = \sum_{\mathbf{a} \in \mathcal{A}} \exp\{\theta \cdot \mathbf{g}(\mathbf{a}, \mathbf{x})\}$
- $\mathcal{A}$: space of possible networks
- θ: model coefficients
- $\mathbf{g}(\mathbf{a}, \mathbf{x})$: sufficient statistics



	1	2	3	4	5	6
1	0	1	0	1	1	0
2	1	0	0	0	1	0
3	0	0	0	0	1	0
4	1	0	0	0	1	0
5	1	1	1	1	0	0
6	0	0	0	0	0	0

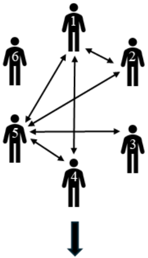
Exponential Family Random Graph Models

Goal: model a network (**A**) given covariates (**X**)

An **ERGM** assumes:

$$\Pr_{\theta}(\mathbf{A} = \mathbf{a} | \mathbf{X} = \mathbf{x}) = \frac{\exp\{\theta \cdot \mathbf{g}(\mathbf{a}, \mathbf{x})\}}{\kappa(\theta, \mathcal{A}, \mathbf{x})}$$

- normalizing constant:
  $\kappa(\theta, \mathcal{A}, \mathbf{x}) = \sum_{\mathbf{a} \in \mathcal{A}} \exp\{\theta \cdot \mathbf{g}(\mathbf{a}, \mathbf{x})\}$
- $\mathcal{A}$: space of possible networks
- θ: model coefficients
- $\mathbf{g}(\mathbf{a}, \mathbf{x})$: sufficient statistics
  - number of edges



	1	2	3	4	5	6
1	0	1	0	1	1	0
2	1	0	0	0	1	0
3	0	0	0	0	1	0
4	1	0	0	0	1	0
5	1	1	1	1	0	0
6	0	0	0	0	0	0

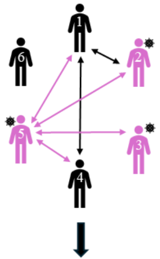
Exponential Family Random Graph Models

Goal: model a network ($\mathbf{A}$) given covariates ($\mathbf{X}$)

An **ERGM** assumes:

$$\Pr_{\theta}(\mathbf{A} = \mathbf{a} | \mathbf{X} = \mathbf{x}) = \frac{\exp\{\theta \cdot \mathbf{g}(\mathbf{a}, \mathbf{x})\}}{\kappa(\theta, \mathcal{A}, \mathbf{x})}$$

- normalizing constant:
  $\kappa(\theta, \mathcal{A}, \mathbf{x}) = \sum_{\mathbf{a} \in \mathcal{A}} \exp\{\theta \cdot \mathbf{g}(\mathbf{a}, \mathbf{x})\}$
- $\mathcal{A}$: space of possible networks
- θ: model coefficients
- $\mathbf{g}(\mathbf{a}, \mathbf{x})$: sufficient statistics
  - number of edges
  - number of edges touching treated students



	1	2	3	4	5	6
1	0	1	0	1	1	0
2	1	0	0	0	1	0
3	0	0	0	0	1	0
4	1	0	0	0	1	0
5	1	1	1	1	0	0
6	0	0	0	0	0	0

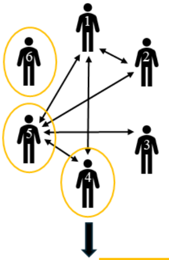
Exponential Family Random Graph Models

Goal: model a network (**A**) given covariates (**X**)

An **ERGM** assumes:

$$\Pr_{\theta}(\mathbf{A} = \mathbf{a} | \mathbf{X} = \mathbf{x}) = \frac{\exp\{\theta \cdot \mathbf{g}(\mathbf{a}, \mathbf{x})\}}{\kappa(\theta, \mathcal{A}, \mathbf{x})}$$

- normalizing constant: $\kappa(\theta, \mathcal{A}, \mathbf{x}) = \sum_{\mathbf{a} \in \mathcal{A}} \exp\{\theta \cdot \mathbf{g}(\mathbf{a}, \mathbf{x})\}$
- $\mathcal{A}$: space of possible networks
- θ: model coefficients
- $\mathbf{g}(\mathbf{a}, \mathbf{x})$: sufficient statistics
  - number of edges
  - number of edges touching treated students
  - number of edges touching infected students



	1	2	3	4	5	6
1	0	1	0	1	1	0
2	1	0	0	0	1	0
3	0	0	0	0	1	0
4	1	0	0	0	1	0
5	1	1	1	1	0	0
6	0	0	0	0	0	0

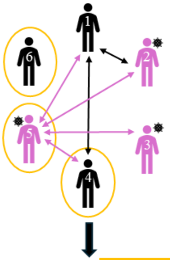
Exponential Family Random Graph Models

Goal: model a network ($\mathbf{A}$) given covariates ($\mathbf{X}$)

An **ERGM** assumes:

$$\Pr_{\theta}(\mathbf{A} = \mathbf{a} | \mathbf{X} = \mathbf{x}) = \frac{\exp\{\theta \cdot \mathbf{g}(\mathbf{a}, \mathbf{x})\}}{\kappa(\theta, \mathcal{A}, \mathbf{x})}$$

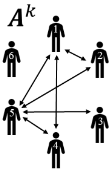
- normalizing constant:
  $\kappa(\theta, \mathcal{A}, \mathbf{x}) = \sum_{\mathbf{a} \in \mathcal{A}} \exp\{\theta \cdot \mathbf{g}(\mathbf{a}, \mathbf{x})\}$
- $\mathcal{A}$: space of possible networks
- θ: model coefficients
- $\mathbf{g}(\mathbf{a}, \mathbf{x})$: sufficient statistics
  - number of edges
  - number of edges touching treated students
  - number of edges touching infected students
  - number of edges touching treated $\times$ infected students



	1	2	3	4	5	6
1	0	1	0	1	1	0
2	1	0	0	0	1	0
3	0	0	0	0	1	0
4	1	0	0	0	1	0
5	1	1	1	1	0	0
6	0	0	0	0	0	0

Separable Temporal Exponential Family Random Graph Models

A **STERGM** assumes that, at each time step $k = 1, \ldots, \tau - 1$:

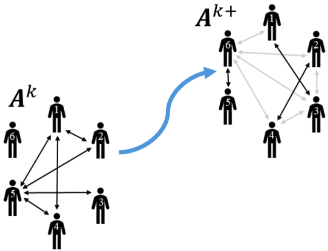


Separable Temporal Exponential Family Random Graph Models

A **STERGM** assumes that, at each time step $k = 1, \dots, \tau - 1$:

- New edges form according to a **formation** ERGM

$$\Pr_{\theta^+}(\mathbf{A}^{k+} = \mathbf{a}^{k+} | \mathbf{A}^k = \mathbf{a}^k, \mathbf{X} = \mathbf{x}) = \frac{\exp\{\theta^+ \cdot \mathbf{g}^+(\mathbf{a}^{k+}, \mathbf{x})\}}{\kappa\{\theta^+, \mathcal{A}^+(\mathbf{a}^k), \mathbf{x}\}}$$

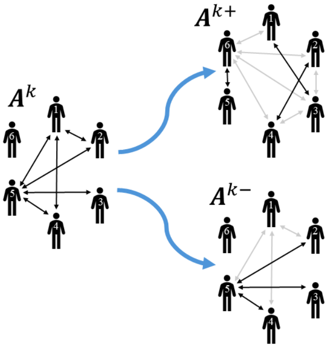


Separable Temporal Exponential Family Random Graph Models

A **STERGM** assumes that, at each time step $k = 1, \ldots, \tau - 1$:

- New edges form according to a **formation** ERGM
- Old edges persist according to a **persistence** ERGM

$$\Pr_{\theta^-}(\mathbf{A}^{k-} = \mathbf{a}^{k-} | \mathbf{A}^k = \mathbf{a}^k, \mathbf{X} = \mathbf{x}) = \frac{\exp\{\theta^- \cdot \mathbf{g}^-(\mathbf{a}^{k-}, \mathbf{x})\}}{\kappa\{\theta^-, \mathcal{A}^-(\mathbf{a}^k), \mathbf{x}\}}.$$

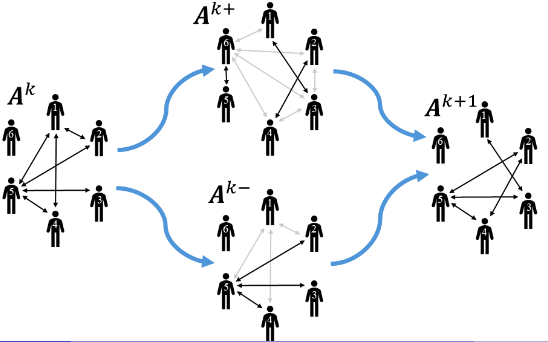


Separable Temporal Exponential Family Random Graph Models

A **STERGM** assumes that, at each time step $k = 1, \ldots, \tau - 1$:

- New edges form according to a **formation** ERGM
- Old edges persist according to a **persistence** ERGM
- The network at time step $k+1$ is the result of formation and persistence

$$\mathbf{A}^{k+1} = \underbrace{\mathbf{A}^{k}}_{\text{previous network}} \cup \underbrace{(\mathbf{A}^{k+} - \mathbf{A}^{k})}_{\text{new edges formed}} - \underbrace{(\mathbf{A}^{k} - \mathbf{A}^{k-})}_{\text{old edges not persisting}}$$



Simulation Study

Goal: use simulated data to investigate the performance of our testing procedure

Simulation Study

Goal: use simulated data to investigate the performance of our testing procedure

- $n \in \{24, 48\}$ students
- $\tau \in \{5, 10\}$ weeks
- 50:50 cluster randomization with 12 equal-sized clusters

Simulation Study

Goal: use simulated data to investigate the performance of our testing procedure

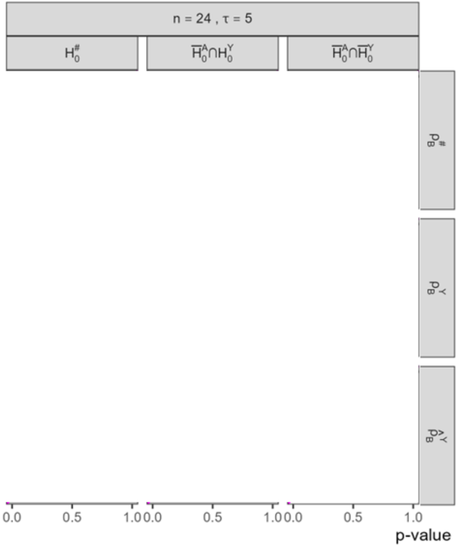
- $n \in \{24, 48\}$ students
- $\tau \in \{5, 10\}$ weeks
- 50:50 cluster randomization with 12 equal-sized clusters
- Three scenarios:
  1. $H_0^\sharp$: no effect of intervention on networks or infection
  2. $\overline{H}_0^A \cap H_0^Y$: no effect of intervention on infection
  3. $\overline{H}_0^A \cap \overline{H}_0^Y$: effect of intervention on both networks and infection

Simulation Study

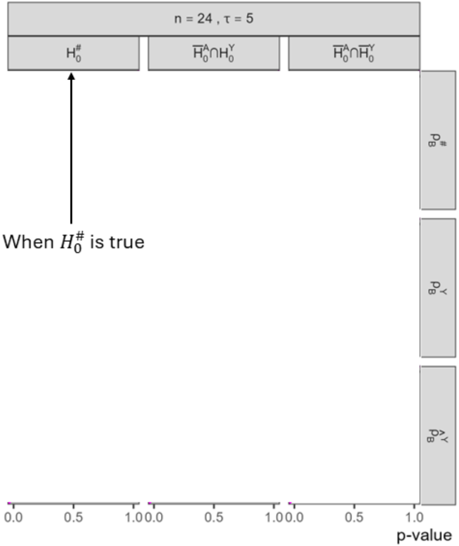
Goal: use simulated data to investigate the performance of our testing procedure

- $n \in \{24, 48\}$ students
- $\tau \in \{5, 10\}$ weeks
- 50:50 cluster randomization with 12 equal-sized clusters
- Three scenarios:
  1. $H_0^\sharp$: no effect of intervention on networks or infection
  2. $\overline{H}_0^A \cap H_0^Y$: no effect of intervention on infection
  3. $\overline{H}_0^A \cap \overline{H}_0^Y$: effect of intervention on both networks and infection
- Three testing procedures:
  1. $\rho_B^\sharp$: testing $H_0^\sharp$
  2. ρ_B^Y: testing H_0^Y using known q
  3. $\widehat{\rho}_B^Y$: testing H_0^Y using estimated q

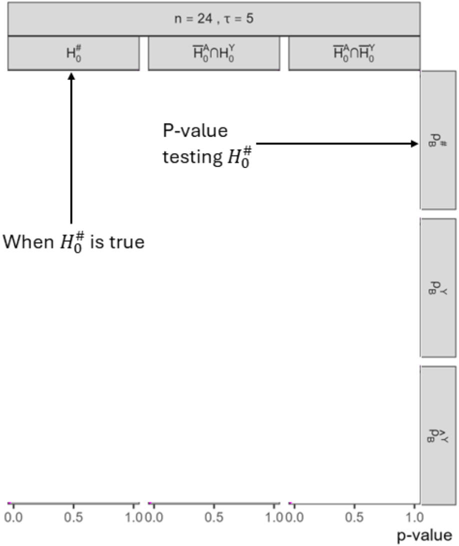
Simulation Results: Empirical CDF of P-Values



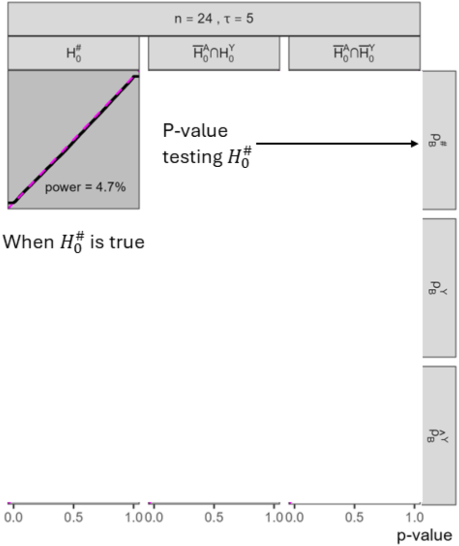
Simulation Results: Empirical CDF of P-Values



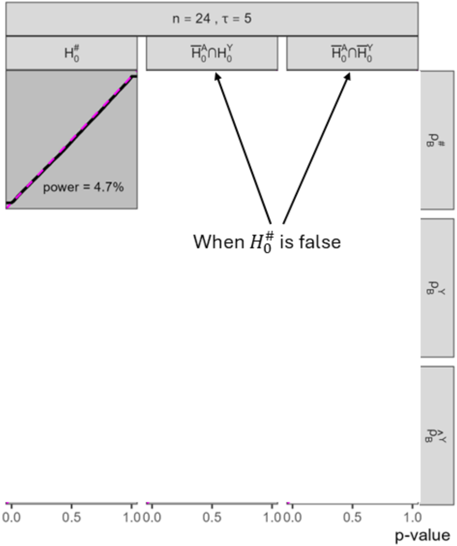
Simulation Results: Empirical CDF of P-Values



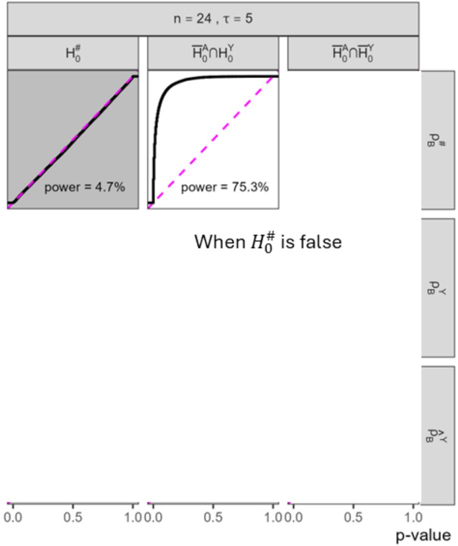
Simulation Results: Empirical CDF of P-Values



Simulation Results: Empirical CDF of P-Values

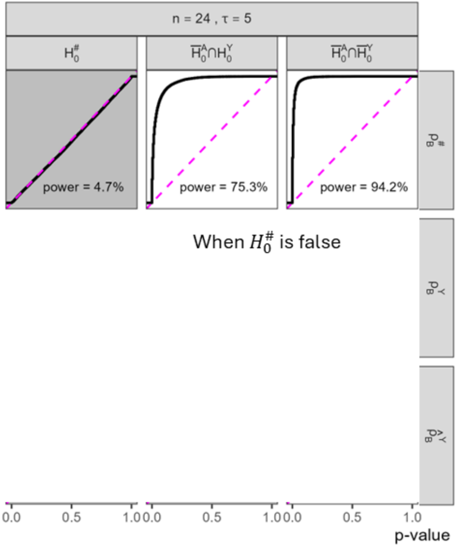


Simulation Results: Empirical CDF of P-Values

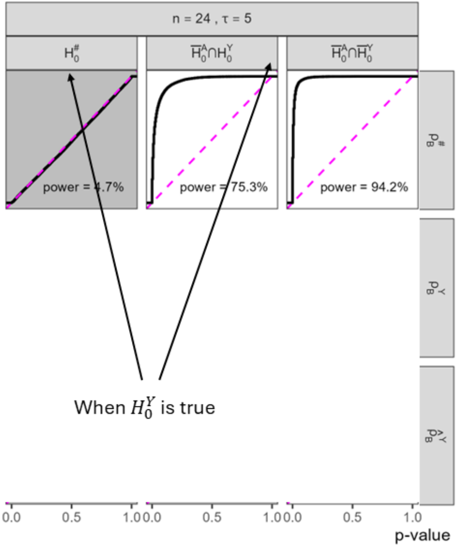


When $H_0^{\#}$ is false

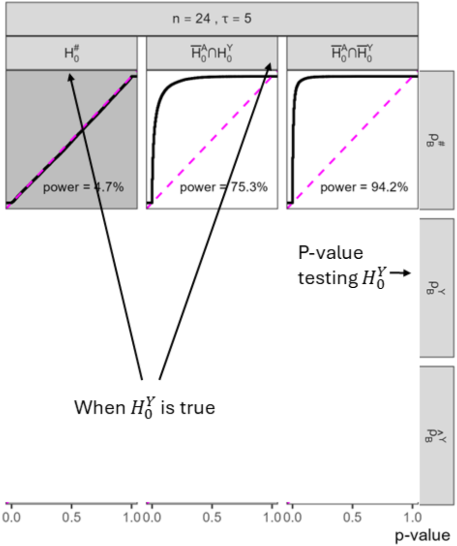
Simulation Results: Empirical CDF of P-Values



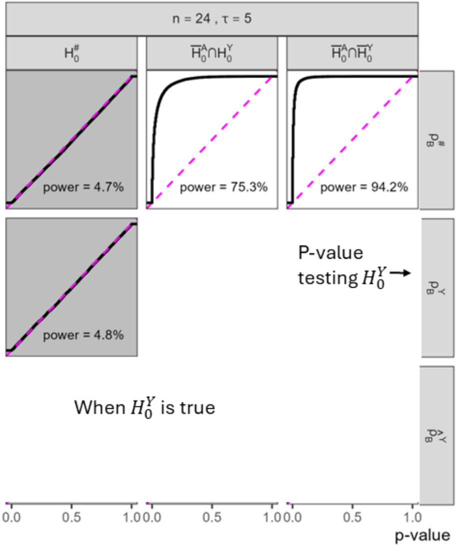
Simulation Results: Empirical CDF of P-Values



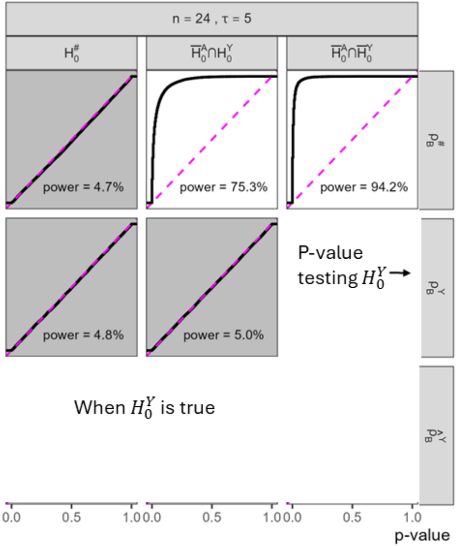
Simulation Results: Empirical CDF of P-Values



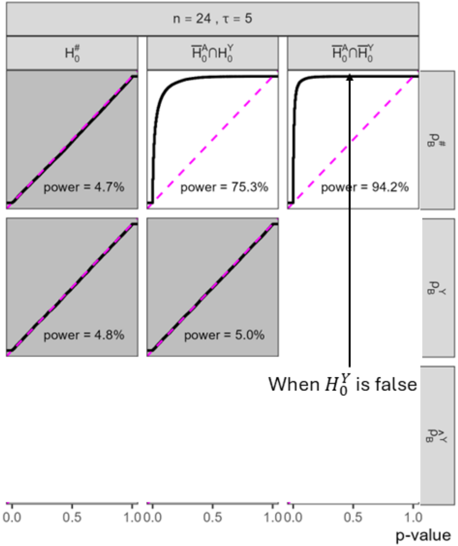
Simulation Results: Empirical CDF of P-Values



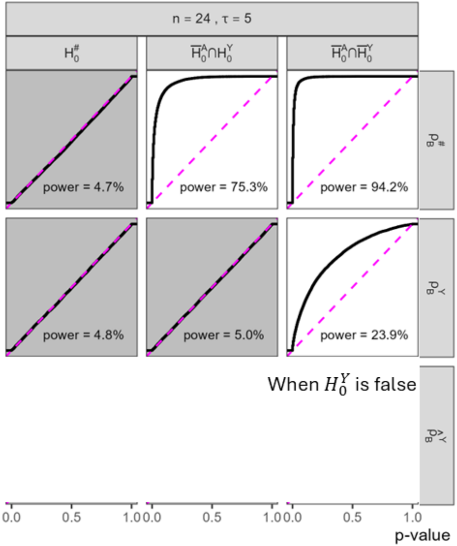
Simulation Results: Empirical CDF of P-Values



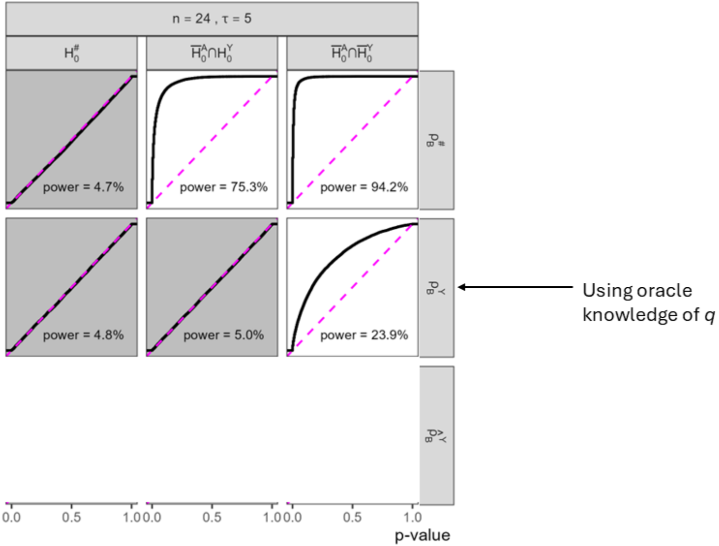
Simulation Results: Empirical CDF of P-Values



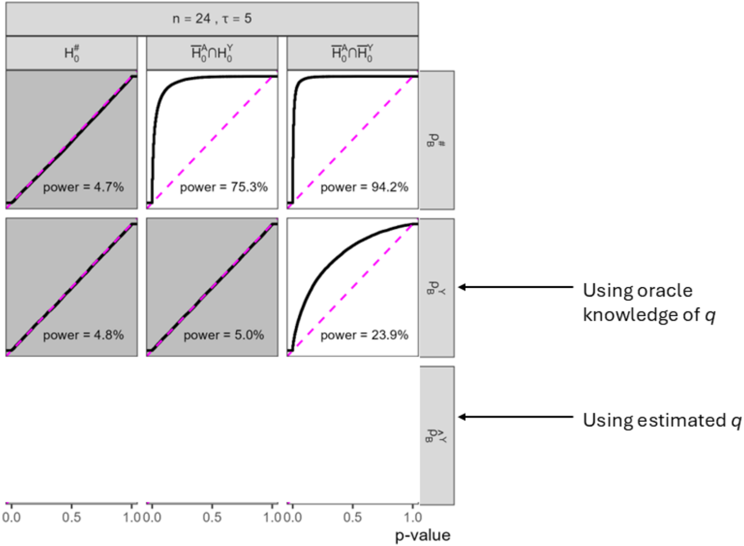
Simulation Results: Empirical CDF of P-Values



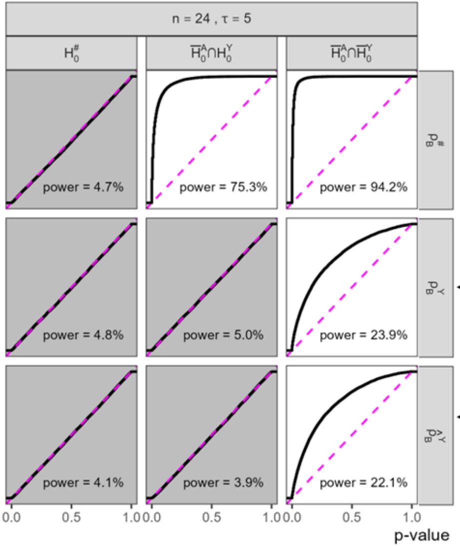
Simulation Results: Empirical CDF of P-Values



Simulation Results: Empirical CDF of P-Values



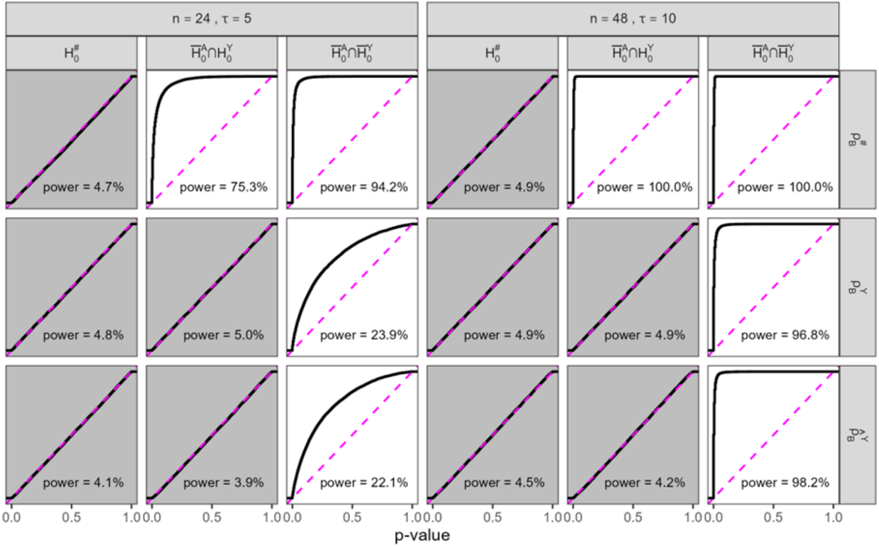
Simulation Results: Empirical CDF of P-Values



Using oracle knowledge of q

Using estimated q

Simulation Results: Empirical CDF of P-Values

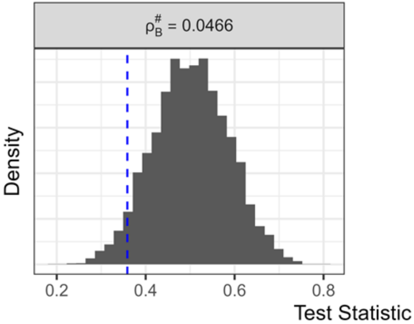


eX-FLU: Hypothesis Test Results

- Test statistic (proportion of possible transmission events attributable to students in the intervention group): $T = 0.359$.

eX-FLU: Hypothesis Test Results

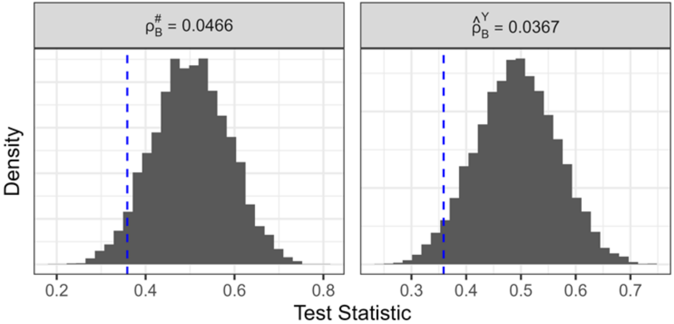
- Test statistic (proportion of possible transmission events attributable to students in the intervention group): $T = 0.359$.



- Encouragement to isolate affects the social network and/or influenze-like illness ($\rho_B^\sharp = 0.0466$)

eX-FLU: Hypothesis Test Results

- Test statistic (proportion of possible transmission events attributable to students in the intervention group): $T = 0.359$.



- Encouragement to isolate affects the social network and/or influenze-like illness ($\rho_B^\sharp = 0.0466$)
- Encouragement to isolate specifically affects influenze-like illness ($\widehat{\rho}_B^Y = 0.0367$)

Future Directions

1. Account for **measurement error** in the self-reported social networks

Future Directions

1. Account for **measurement error** in the self-reported social networks
   - Using mismeasured networks does not hurt the type I error, but may hurt power

Future Directions

1. Account for **measurement error** in the self-reported social networks
   - Using mismeasured networks does not hurt the type I error, but may hurt power
   - Use Bluetooth data from eX-FLU sub-study

Future Directions

1. Account for **measurement error** in the self-reported social networks
   - Using mismeasured networks does not hurt the type I error, but may hurt power
   - Use Bluetooth data from eX-FLU sub-study
2. Focus on **estimation** as opposed to **hypothesis testing**

Future Directions

1. Account for **measurement error** in the self-reported social networks
   - Using mismeasured networks does not hurt the type I error, but may hurt power
   - Use Bluetooth data from eX-FLU sub-study
2. Focus on **estimation** as opposed to **hypothesis testing**
   - trade-off between robustness and utility

Future Directions

1. Account for **measurement error** in the self-reported social networks
   - Using mismeasured networks does not hurt the type I error, but may hurt power
   - Use Bluetooth data from eX-FLU sub-study
2. Focus on **estimation** as opposed to **hypothesis testing**
   - trade-off between robustness and utility
3. **Design future trials** with independent clusters
   - can allow **identification** of more causal estimands

Thank you!

Questions?

References I

Allison E. Aiello, Amanda M. Simanek, Marisa C. Eisenberg, Alison R. Walsh, Brian Davis, Erik Volz, Caroline Cheng, Jeanette J. Rainey, Amra Uzicanin, Hongjiang Gao, Nathaniel Osgood, Dylan Knowles, Kevin Stanley, Kara Tarter, and Arnold S. Monto. Design and methods of a social network isolation study for reducing respiratory infection transmission: The eX-FLU cluster randomized trial. *Epidemics*, 15:38–55, June 2016. ISSN 17554365. doi: 10.1016/j.epidem.2016.01.001. URL https://linkinghub.elsevier.com/retrieve/pii/S1755436516000025.

Shaina J. Alexandria, Michael G. Hudgens, and Allison E. Aiello. Assessing Intervention Effects in a Randomized Trial Within a Social Network. *Biometrics*, 79(2):1409–1419, June 2023. ISSN 0006-341X, 1541-0420. doi: 10.1111/biom.13606. URL https://academic.oup.com/biometrics/article/79/2/1409-1419/7513977.

R. A. Fisher. The design of experiments. *Nature*, 137(3459):252–254, February 1936. ISSN 0028-0836, 1476-4687. doi: 10.1038/137252a0.

Michael G Hudgens and M. Elizabeth Halloran. Toward causal inference with interference. *Journal of the American Statistical Association*, 103(482):832–842, 2008. ISSN 0162-1459, 1537-274X. doi: 10.1198/016214508000000292.

David M Ritzwoller, Joseph P Romano, and Azeem M Shaikh. Randomization inference: Theory and applications. 2024.

Yao Zhang and Qingyuan Zhao. What is a randomization test? *Journal of the American Statistical Association*, 118(544):2928–2942, October 2023. ISSN 0162-1459, 1537-274X. doi: 10.1080/01621459.2023.2199814.

P.N. Zivich, M.C. Eisenberg, A.S. Monto, A. Uzicanin, R. S. Baric, T. P. Sheahan, J. J. Rainey, H. Gao, and A. E. Aiello. Transmission of viral pathogens in a social network of university students: the eX-FLU study. *Epidemiology and Infection*, 148:e267, 2020. ISSN 0950-2688, 1469-4409. doi: 10.1017/S0950268820001806. URL https://www.cambridge.org/core/product/identifier/S0950268820001806/type/journal_article.

Choice of Test Statistic

Test statistics: "proportion of possible transmission events attributable to students in the intervention group:"

$$T^{**}(\mathbf{Z}, \overline{\mathbf{A}}, \overline{\mathbf{Y}}) = \frac{\sum_{k=2}^{\tau} \sum_{i=1}^{n} \sum_{j>i} Z_i E_{ijk}^{**}}{\sum_{k=2}^{\tau} \sum_{i=1}^{n} \sum_{j>i} E_{ijk}^{**}}.$$

	From Infected to Infected	**From Infected**
Contact at $k-1$	$E_{ijk}^{11} = Y_i^{k-1} A_{ij}^{k-1} Y_j^k$	$E_{ijk}^{12} = Y_i^{k-1} A_{ij}^{k-1}$
Contact at k	$E_{ijk}^{21} = Y_i^{k-1} A_{ij}^{k} Y_j^k$	$E_{ijk}^{22} = Y_i^{k-1} A_{ij}^{k}$
Contact at $k-1$ **or** k	$E_{ijk}^{31} = Y_i^{k-1} (A_{ij}^{k-1} \vee A_{ij}^{k}) Y_j^k$	$E_{ijk}^{32} = Y_i^{k-1} (A_{ij}^{k-1} \vee A_{ij}^{k})$
Contact at $k-1$ **and** k	$E_{ijk}^{41} = Y_i^{k-1} (A_{ij}^{k-1} * A_{ij}^{k}) Y_j^k$	$E_{ijk}^{42} = Y_i^{k-1} (A_{ij}^{k-1} * A_{ij}^{k})$

Table: Definitions of a possible transmission event E_{ijk} from student i to student j at time k. Y_i^k is an indicator for student i being infected at week k, A_{ij}^k is an indicator for students i and j being in contact at week k, and $a \vee b$ denotes the maximum of a and b.

ERGM Model Formulation

$$\Pr_\theta(\mathbf{A} = \mathbf{a} | \mathbf{X} = \mathbf{x}) = \frac{\exp\{\theta \cdot \mathbf{g}(\mathbf{a}, \mathbf{x})\}}{\kappa(\theta, \mathcal{A}, \mathbf{x})},$$

$$\kappa(\theta, \mathcal{A}, \mathbf{x}) = \sum_{\mathbf{a} \in \mathcal{A}} \exp\{\theta \cdot \mathbf{g}(\mathbf{a}, \mathbf{x})\}$$

For example, in the simulation study and eX-FLU application,

$$\mathbf{g}(\mathbf{a}, \mathbf{x}) = \begin{bmatrix} \text{\# edges} \\ \text{\# edges touching a treated node} \\ \text{\# edges touching an infected node} \\ \text{\# edges touching a treated and infected node} \\ \text{\# edges between roommate pairs} \end{bmatrix} = \begin{bmatrix} \sum_{i,j} a_{ij} \\ \sum_{i,j} a_{ij}(z_i + z_j) \\ \sum_{i,j} a_{ij}(y_i + y_j) \\ \sum_{i,j} a_{ij}(z_i y_i + z_j y_j) \\ \sum_{i,j} a_{ij} \mathbb{1}(i, j \text{ roomates}) \end{bmatrix}$$

ERGM Change Statistic Model Formulation

- **Change statistic**: $\delta_{\mathbf{g}}(\mathbf{a},\mathbf{x})_{ij} = \mathbf{g}(\mathbf{a}_{ij}^{+},\mathbf{x}) - \mathbf{g}(\mathbf{a}_{ij}^{-},\mathbf{x})$ is the change in network statistic that would occur if a_{ij} were changed from 0 to 1
  - where $\mathbf{a}_{ij}^{+}$ and $\mathbf{a}_{ij}^{-}$ represent the network $\mathbf{a}$ with dyad a_{ij} set to 1 or 0, respectively
- Then the equivalent ERGM specification is

$$\text{logit}\{\Pr_{\theta}(A_{ij} = 1 | \mathbf{A}_{ij}^{C} = \mathbf{a}_{ij}^{C}, \mathbf{X} = \mathbf{x})\} = \theta^{k}\delta_{\mathbf{g}}(\mathbf{a},\mathbf{x})_{ij}$$

  - where $\mathbf{A}_{ij}^{C}$ represents all dyads in $\mathbf{A}$ except A_{ij}
- **Interpretation of θ**: the change in conditional log-odds of the network associated with a one-unit increase in the corresponding component of $\mathbf{g}(\mathbf{a},\mathbf{x})$ resulting from switching a particular dyad A_{ij} from 0 to 1 and leaving the rest of the network fixed at $\mathbf{A}_{ij}^{C}$

Dyadic Independence EGRMs

- **Dyadic independence term**: a component g of $\mathbf{g}$ in an ERGM for which the corresponding change statistic $\delta_g(\mathbf{a}, \mathbf{x})_{ij}$ can be calculated for any i, j without knowing $\mathbf{a}$
  - For example, if $g(\mathbf{a}, \mathbf{x}) = \sum_{i,j} a_{ij}(Z_i + Z_j)$ counts the number of edges touching treated nodes, then $\delta_g(\mathbf{a}, \mathbf{x})_{ij} = z_i + z_j$ doesn't depend on $\mathbf{a}$
- **Dyadic independence ERGM**: an ERGM with only dyadic independence terms
  - replace $\delta_g(\mathbf{a}, \mathbf{x})_{ij}$ with $\delta_g(\mathbf{x})_{ij}$ and write the model as

$$\text{logit}\{\Pr\big(A_{ij} = 1 | \mathbf{X} = \mathbf{x}\big)\} = \theta \cdot \delta_g(\mathbf{x})_{ij}$$

- **Interpretation of θ**: the change in log-odds of the network associated with a one-unit increase in the corresponding component of $\mathbf{g}(\mathbf{a}, \mathbf{x})$ resulting from switching a particular dyad A_{ij} from 0 to 1

STERGM Model Formulation

- **Formation model** is an ERGM conditional on only adding edges:

$$\Pr_{\theta^+}(\mathbf{A}^{k+} = \mathbf{a}^{k+} | \mathbf{A}^k = \mathbf{a}^k, \mathbf{X} = \mathbf{x}) = \frac{\exp\{\theta^+ \cdot \mathbf{g}^+(\mathbf{a}^{k+}, \mathbf{x})\}}{\kappa\{\theta^+, \mathcal{A}^+(\mathbf{a}^k), \mathbf{x}\}}$$

  - $\mathcal{A}^+(\mathbf{a})$: space of possible networks that can be formed by adding edges to $\mathbf{a}$

- **Persistence model** is an ERGM conditional on only removing edges:

$$\Pr_{\theta^-}(\mathbf{A}^{k-} = \mathbf{a}^{k-} | \mathbf{A}^k = \mathbf{a}^k, \mathbf{X} = \mathbf{x}) = \frac{\exp\{\theta^- \cdot \mathbf{g}^-(\mathbf{a}^{k-}, \mathbf{x})\}}{\kappa\{\theta^-, \mathcal{A}^-(\mathbf{a}^k), \mathbf{x}\}}$$

  - $\mathcal{A}^-(\mathbf{a})$: space of possible networks that can be formed by removing edges to $\mathbf{a}$

- A **STERGM** assumes the network at time $k+1$ is then the result of applying the changes in $\mathbf{A}^{k+}$ and $\mathbf{A}^{k-}$ to $\mathbf{A}^k$:

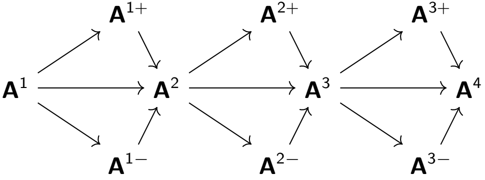
$$\mathbf{A}^{k+1} = \underbrace{\mathbf{A}^k}_{\text{previous network}} \cup \underbrace{(\mathbf{A}^{k+} - \mathbf{A}^k)}_{\text{new edges formed}} - \underbrace{(\mathbf{A}^k - \mathbf{A}^{k-})}_{\text{old edges not persisting}}$$

Separability of STERGMs

1. $\mathbf{A}^{k+} \perp\!\!\!\perp \mathbf{A}^{k-} | \mathbf{A}^{k}$, i.e., the formation and persistence processes are conditionally independent give the network at time k

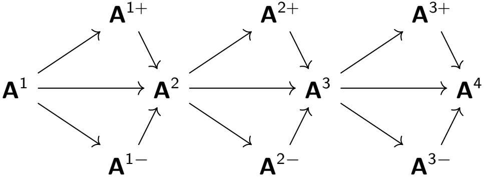
Separability of STERGMs

1. $\mathbf{A}^{k+} \perp\!\!\!\perp \mathbf{A}^{k-} | \mathbf{A}^{k}$, i.e., the formation and persistence processes are conditionally independent give the network at time k



Separability of STERGMs

1. $\mathbf{A}^{k+} \perp\!\!\!\perp \mathbf{A}^{k-} | \mathbf{A}^{k}$, i.e., the formation and persistence processes are conditionally independent give the network at time k
2. the parameter space for $\theta = (\theta^+, \theta^-)$ is the product of the parameter spaces for θ^+ and θ^-



Simulation Setup

- $n \in \{24, 48\}$ students were equally divided into two residence halls, each with six clusters of equal size
- Within each residence hall group, the clusters were randomized using a 50:50 allocation to either the intervention group or the control group
- Five pairs of students were randomly selected to be roommates, meaning they had contact with each other each week
- Baseline $(k = 1)$ social contacts were simulated between each pair of students with probability 0.5
- Baseline infection statuses were simulated for each student with probability 0.5
- Social networks were simulated over the remaining $\tau \in \{5, 10\}$ weeks according to a STERGM with both formation and persistence models including edge count, intervention assignment Z, infection status Y, $Z \times Y$ interaction, and an offset to force constant edges between roommates
- Infection statuses were simulated for each student i at each week $k \in \{2, \ldots, \tau\}$ with probability $\Pr\left(Y_i^k = 1 | \overline{\mathbf{A}}^{k-1}, \overline{\mathbf{Y}}^{k-1}\right) = g(\sum_{j=1}^n A_{ij}^{k-1} Y_j^{k-1})$, where $\sum_{j=1}^n A_{ij}^{k-1} Y_j^{k-1}$ is the number of infected contacts at the previous week, and $g : [0, \infty) \to [0, 1]$ is a non-decreasing function

Simulation Setup

Three scenarios:

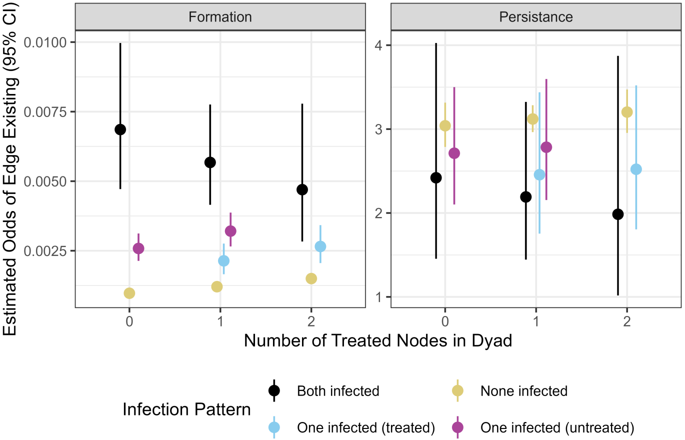
Null Hypothesis	Formation Model Parameters θ^+	Persistence Model Parameters θ^-	Infection Probability Function g
$H_0^\sharp$	$(-0.5, 0, 0, 0)$	$(-0.5, 0, 0, 0)$	$g(s) = 0.5$
$\overline{H}_0^A \cap H_0^Y$	$(-0.2, 0, 0, -1)$	$(-0.2, 0, 0, -1)$	$g(s) = 0.5$
$\overline{H}_0^A \cap \overline{H}_0^Y$	$(-0.2, 0, 0, -1)$	$(-0.2, 0, 0, -1)$	$g(s)$ increasing in s

Table: Data generating process for the simulation study. The null hypothesis refers to the null hypothesis that is true under the data generating process, formation model parameters are $\theta^+ = (\theta^+_{\text{edges}}, \theta^+_Z, \theta^+_Y, \theta^+_{ZY})$, persistence model parameters are $\theta^- = (\theta^-_{\text{edges}}, \theta^-_Z, \theta^-_Y, \theta^-_{ZY})$, and the infection probability function g provides a unit's probability of being infected at week k given their number of infected neighbors at week $k-1$.

Three p-values:

- $\rho_B^\sharp$: testing the sharp null
- ρ_B^Y: testing H_0^Y using known q
- $\widehat{\rho}_B^Y$: testing H_0^Y using estimated q

eX-FLU: STERGM Results



Type I Error Control

- **Proposition 2.1**: Let T_N be a test statistic with CDF F_N under H_0.
  1. Under H_0, $F_N(T_N)$ stochastically dominates a Uniform$(0,1)$ distribution for any N.
  2. If the test statistic T_N has a continuous limiting distribution, then $F_N(T_N) \to^d$ Uniform$(0,1)$.
- **Corollary 2.2**:
  1. Under $H_0^\sharp$, the sharp null p-value $\rho_N^\sharp$ stochastically dominates a Uniform$(0,1)$ distribution for any N.
  2. Under H_0^Y, the oracle p-value ρ_N^Y stochastically dominates a Uniform$(0,1)$ distribution for any N.
  3. If the test statistic T_N has a continuous limiting distribution, then $\rho_N^\sharp \to^d$ Uniform$(0,1)$ under $H_0^\sharp$ and $\rho_N^Y \to^d$ Uniform$(0,1)$ under H_0^Y.

Type I Error Control

- **Proposition 2.3**: Let $q(\mathbf{a}, \mathbf{z}, \theta) \equiv \Pr\{\mathbf{A}(\mathbf{z}) = \mathbf{a}; \theta\}$ denote the PMF of the distribution of stochastic potential networks $\mathbf{A}(\mathbf{z})$ at parameter value θ, let $\widehat{\theta}_N$ denote the estimator of θ, and let $F_N(\cdot; \theta)$ denote the CDF of the test statistic T_N at θ, with limiting CDF $F(\cdot; \theta)$. Let θ_0 denote the true value of θ. Assume the following:

(A1) $\widehat{\theta}_N \to^p \theta_0$

(A2) $F(t; \theta_0)$ is continuous in t on $\mathbb{R}$

(A3) there exists a $\delta_0 > 0$ such that

$$\sup_{\theta \in B_{\delta_0}(\theta_0)} \sup_{t \in \mathbb{R}} |F_N(t; \theta) - F(t; \theta)| \to 0$$

(A4) $F(t; \theta)$ is continuous in θ at θ_0 uniformly in t, i.e.,

$$\lim_{\theta \to \theta_0} \sup_{t \in \mathbb{R}} |F(t; \theta) - F(t; \theta_0)| = 0$$

Then the plug-in p-value $\widehat{\rho}_N^Y$ converges in distribution to Uniform$(0, 1)$.

Type I Error Control

- **Proposition 2.4**: Let $\rho_N = F_N(T_N)$ for test statistic T_N and CDF F_N (not necessarily the true CDF of T_N). Let $T_N^* = h_N(T_N)$ for a sequence of deterministic, strictly increasing functions h_N. Define $F_N^*(t) = F_N\{h_N^{-1}(t)\}$, the (not necessarily true) CDF of the transformed test statistic, and let $\rho_N^* = F_N^*(T_N^*)$. Then $\rho_N^* = \rho_N$.
- **Corollary 2.5**: Let h_N be a sequence of deterministic, strictly increasing functions.
  1. If the hypotheses of Proposition 2.1 are met for a test statistic T_N, then the results also hold for $T_N^* = h_N(T_N)$.
  2. If the hypotheses of Proposition 2.3 are met for $T_N, F_N(\cdot;\theta)$, then the results also hold for $T_N^* = h_N(T_N), F_N^*(\cdot;\theta) = F_N\{h_N^{-1}(\cdot);\theta\}$.