Causal inference for infectious disease intervention in inter-connected clusters

Xiaoxuan Cai

Joint work with Eben Kenah, Forrest W. Crawford

Department of Statistics
Ohio State University

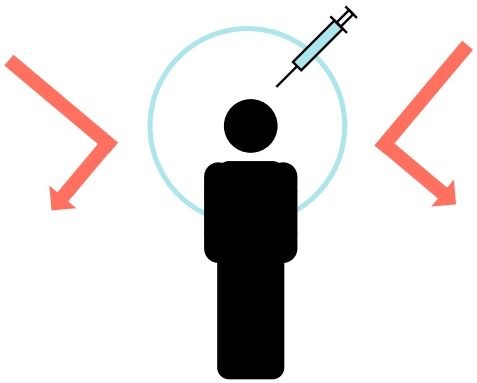
Joint Statistical Meeting 2022

(References and slides are available on my personal website.)

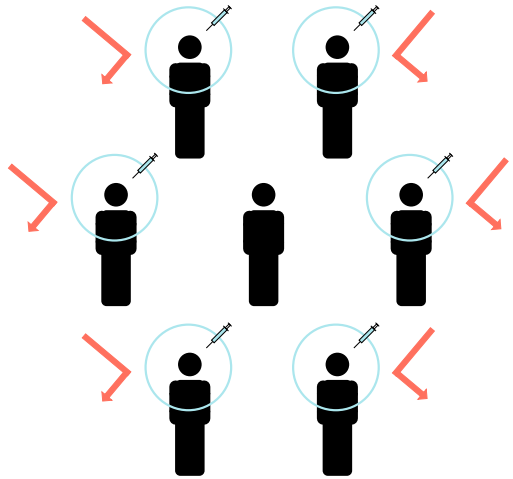
Infectious disease and vaccination

- Direct protection for the treated individuals:
  - direct effect, vaccine efficacy, susceptibility effect, ...
- Indirect protection for the surrounding individuals:
  - indirect effect, herd immunity, contagion effect, infectiousness effect, ...

(Examples: vaccines for Polio, Influenza, HIV/AIDS, Malaria, and etc).



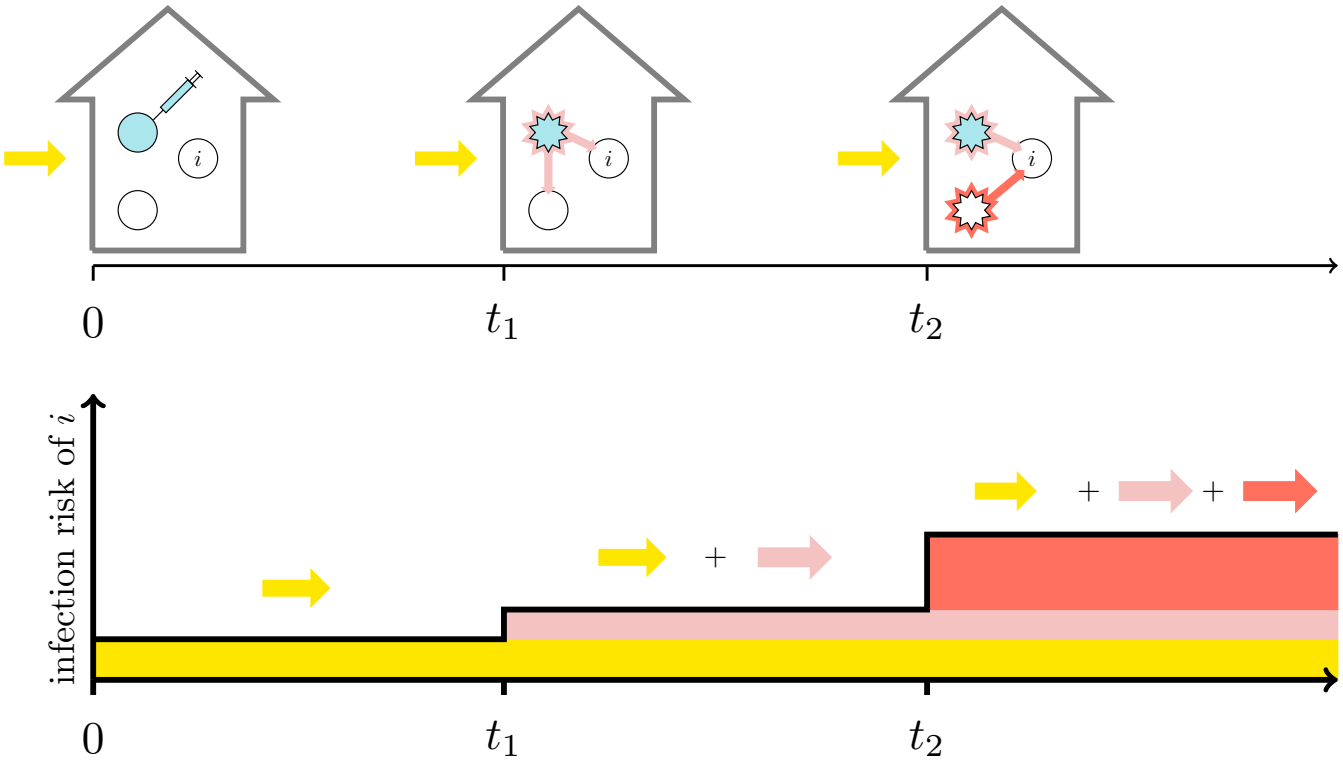
Direct protection

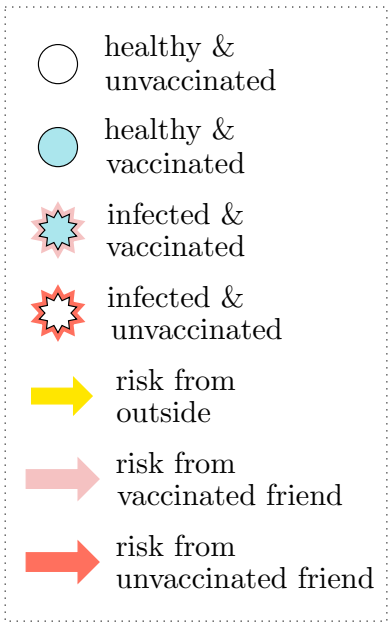


Indirect protection

How epidemiologists understand disease transmission

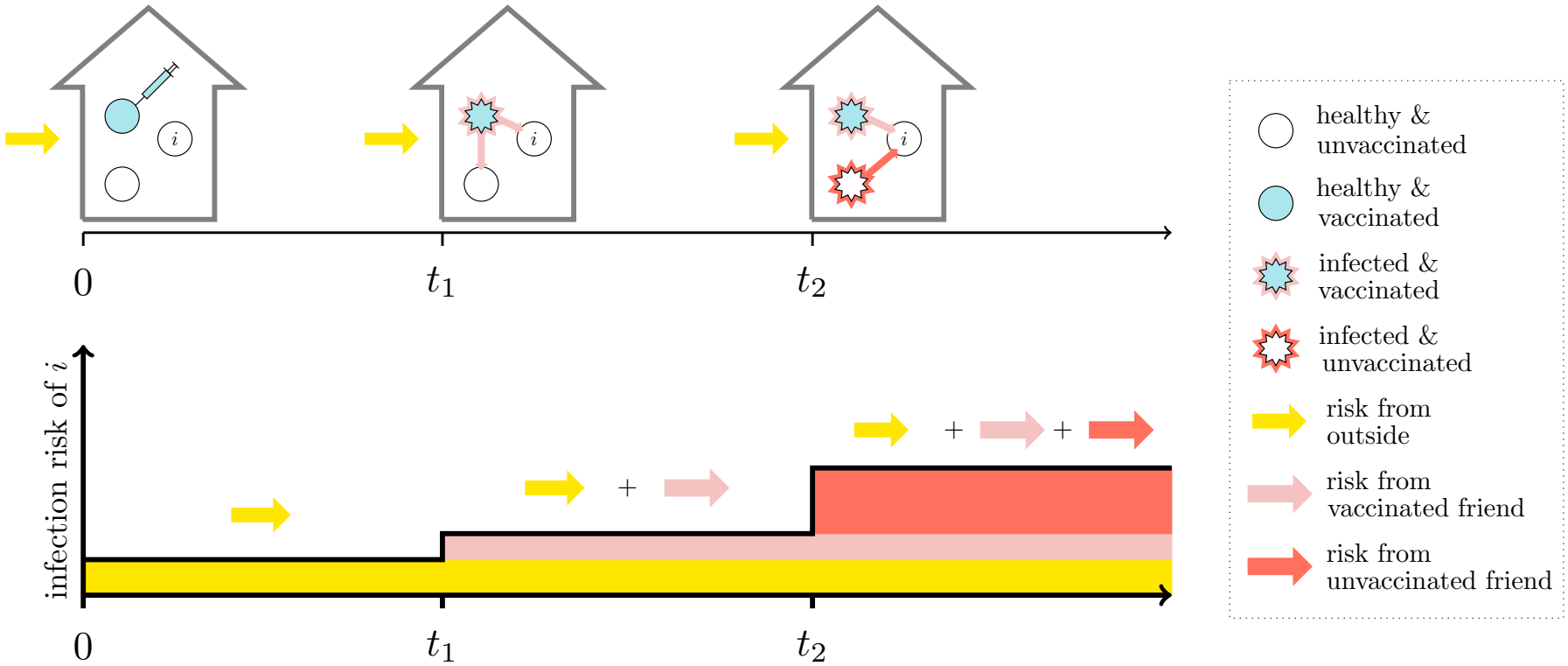
For a focal individual i, the risk of infection increases as more neighbors become infectious and depend on neighbors' vaccination status.





How epidemiologists understand disease transmission

For a focal individual i, the risk of infection increases as more neighbors become infectious and depend on neighbors' vaccination status.

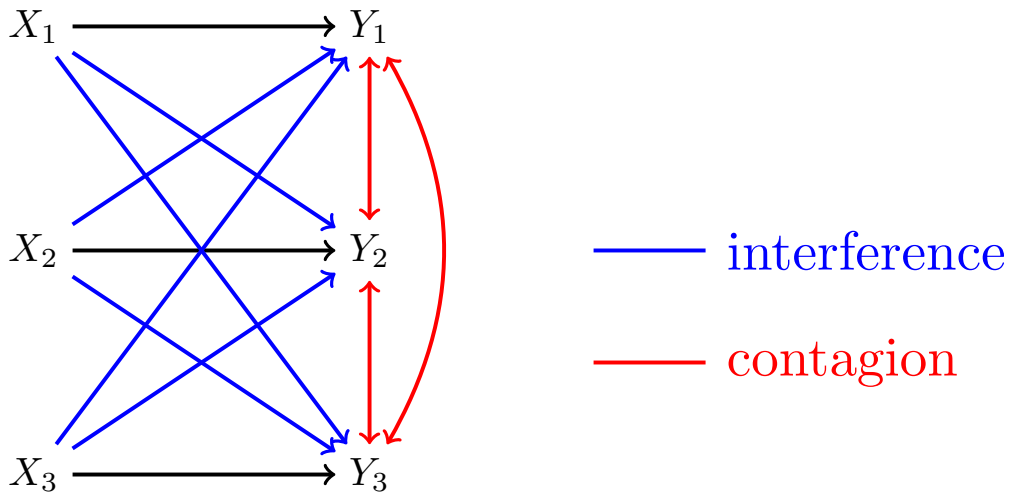


One infection outcome depends on (i) its **own treatment**, (ii) **treatments of others**, and (iii) **infection times of others** (e.g. t_1, t_2).

Why is infectious disease difficult to study in causal inference?

1. Interdependence of outcomes and treatments across subjects

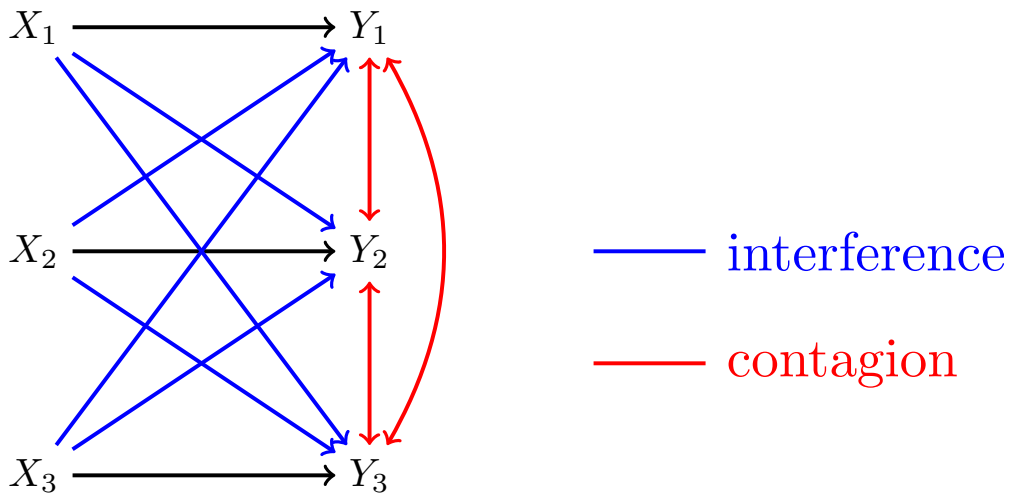
   Consider a interconnected three individuals with treatments (X_1, X_2, X_3) and infection outcomes (Y_1, Y_2, Y_3).



Why is infectious disease difficult to study in causal inference?

1. Interdependence of outcomes and treatments across subjects

   Consider a interconnected three individuals with treatments (X_1, X_2, X_3) and infection outcomes (Y_1, Y_2, Y_3).

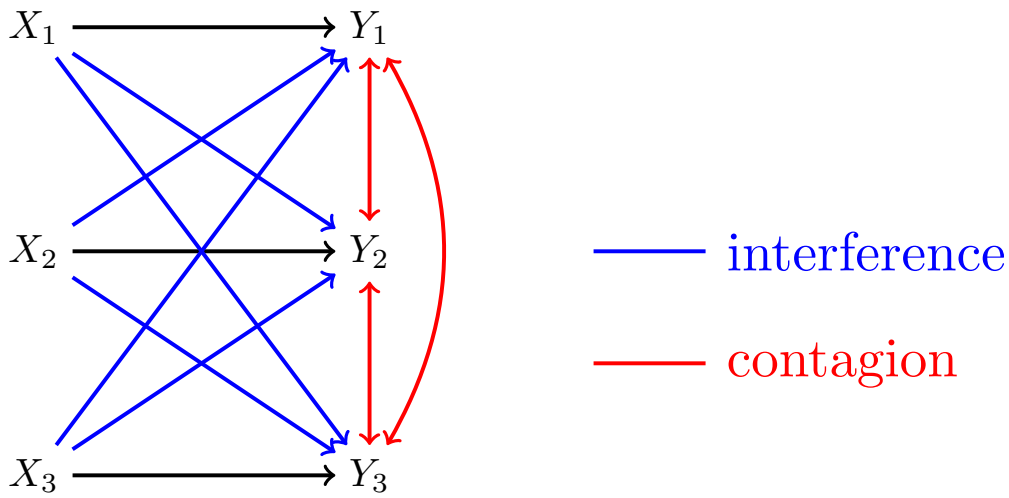


   - The infection outcome of one individual depends on others' treatments
   - The infection outcome of one individual depends on other's outcomes, since it is transmissible.

Why is infectious disease difficult to study in causal inference?

1. Interdependence of outcomes and treatments across subjects

   Consider a interconnected three individuals with treatments (X_1, X_2, X_3) and infection outcomes (Y_1, Y_2, Y_3).



   - The infection outcome of one individual depends on others' treatments
   - The infection outcome of one individual depends on other's outcomes, since it is transmissible.

   Bidirectional arrows causes problems in causal identification

Why is infectious disease difficult to study in causal inference?

1. Interdependence of outcomes and treatments across subjects
2. Stochastic processes of "exposure to infection"
   - "Exposure to infection" is determined by stochastic infection outcomes of others, whose distribution depends on their treatments.

Why is infectious disease difficult to study in causal inference?

1. Interdependence of outcomes and treatments across subjects
2. Stochastic processes of "exposure to infection"

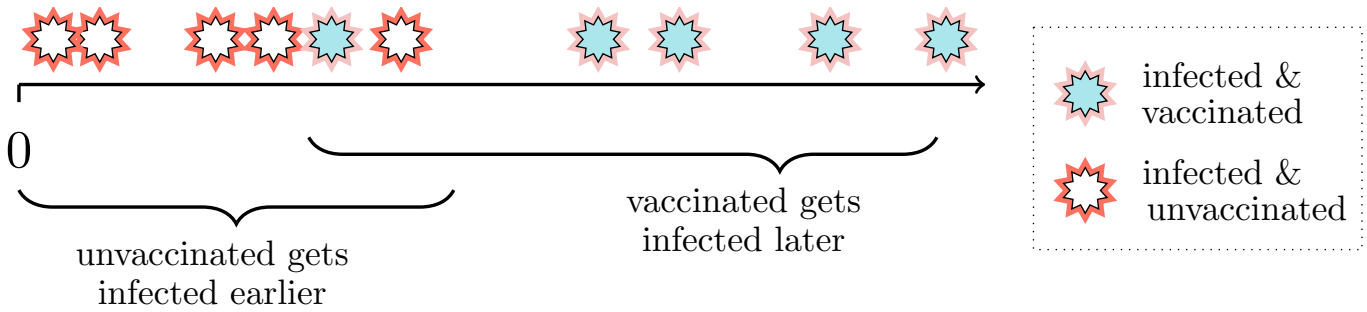
For example, earlier infections of neighbors (smaller t_1 and t_2) or higher infectiousness due to unvaccination (bigger jumps) increases overall "exposure to infection" and consequently infection risk.

Why is infectious disease difficult to study in causal inference?

1. Interdependence of outcomes and treatments across subjects
2. Stochastic processes of "exposure to infection"
3. Bias due to differential "exposure to infection"

Can we directly compare treated and untreated individuals using randomization?

$$E[Y_i|X_i = 1] - E[Y_i|X_i = 0]$$



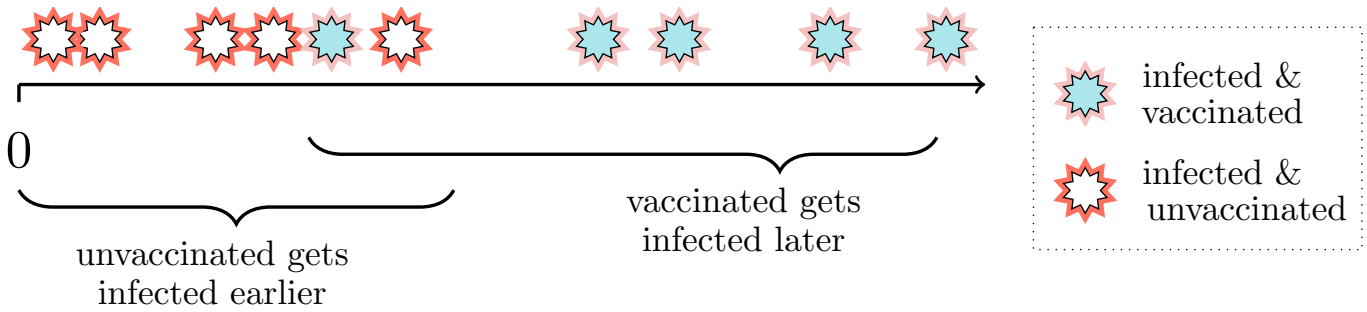
For example, vaccinated individuals endure higher exposure to infection, which is not a fair comparison. $\rightarrow$ Effect is under-estimated!

Why is infectious disease difficult to study in causal inference?

1. Interdependence of outcomes and treatments across subjects
2. Stochastic processes of "exposure to infection"
3. Bias due to differential "exposure to infection"

Can we directly compare treated and untreated individuals using randomization?

$$E[Y_i|X_i = 1] - E[Y_i|X_i = 0]$$

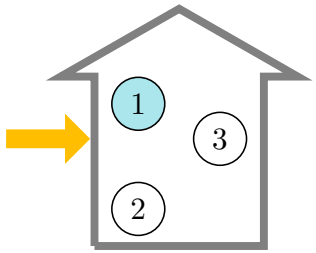


For example, vaccinated individuals endure higher exposure to infection, which is not a fair comparison. $\rightarrow$ Effect is under-estimated!

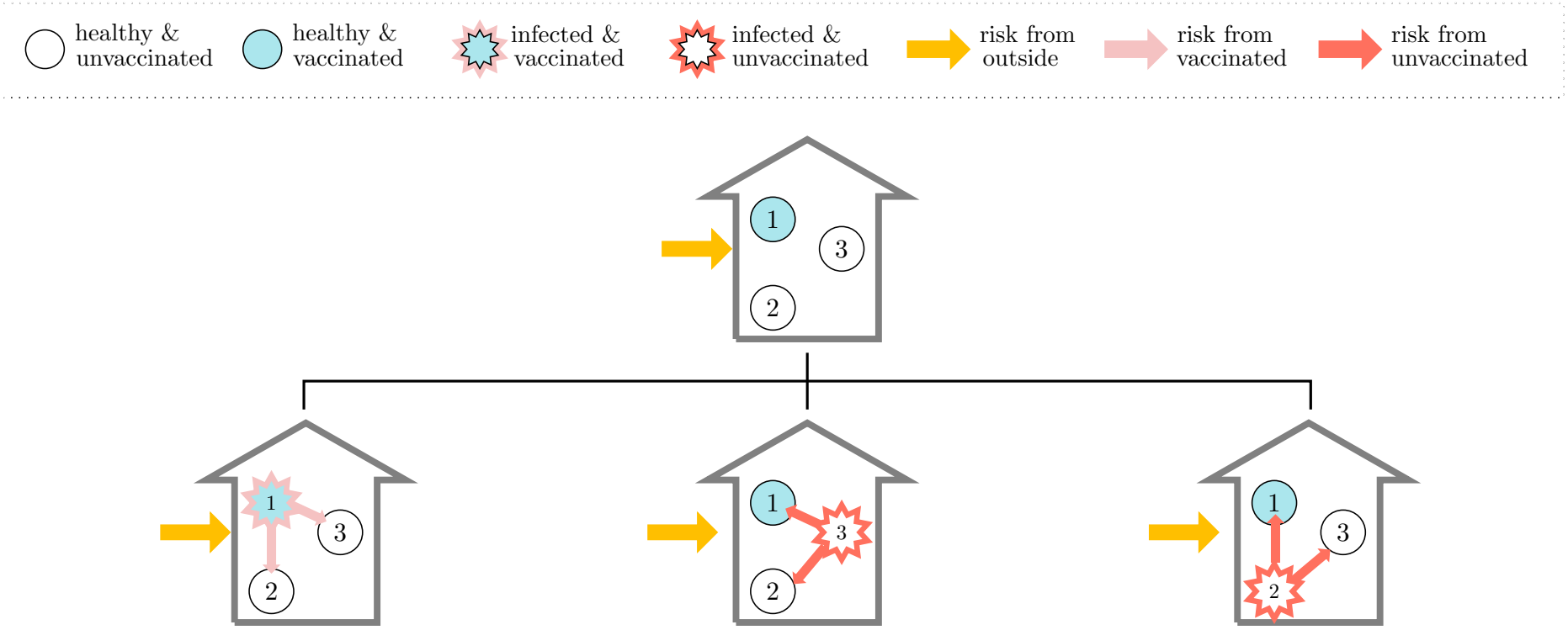
NO! It may be biased due to differential "exposure to infection". [1,2,3].

So how to solve the causal identification problem for infectious disease outcomes?

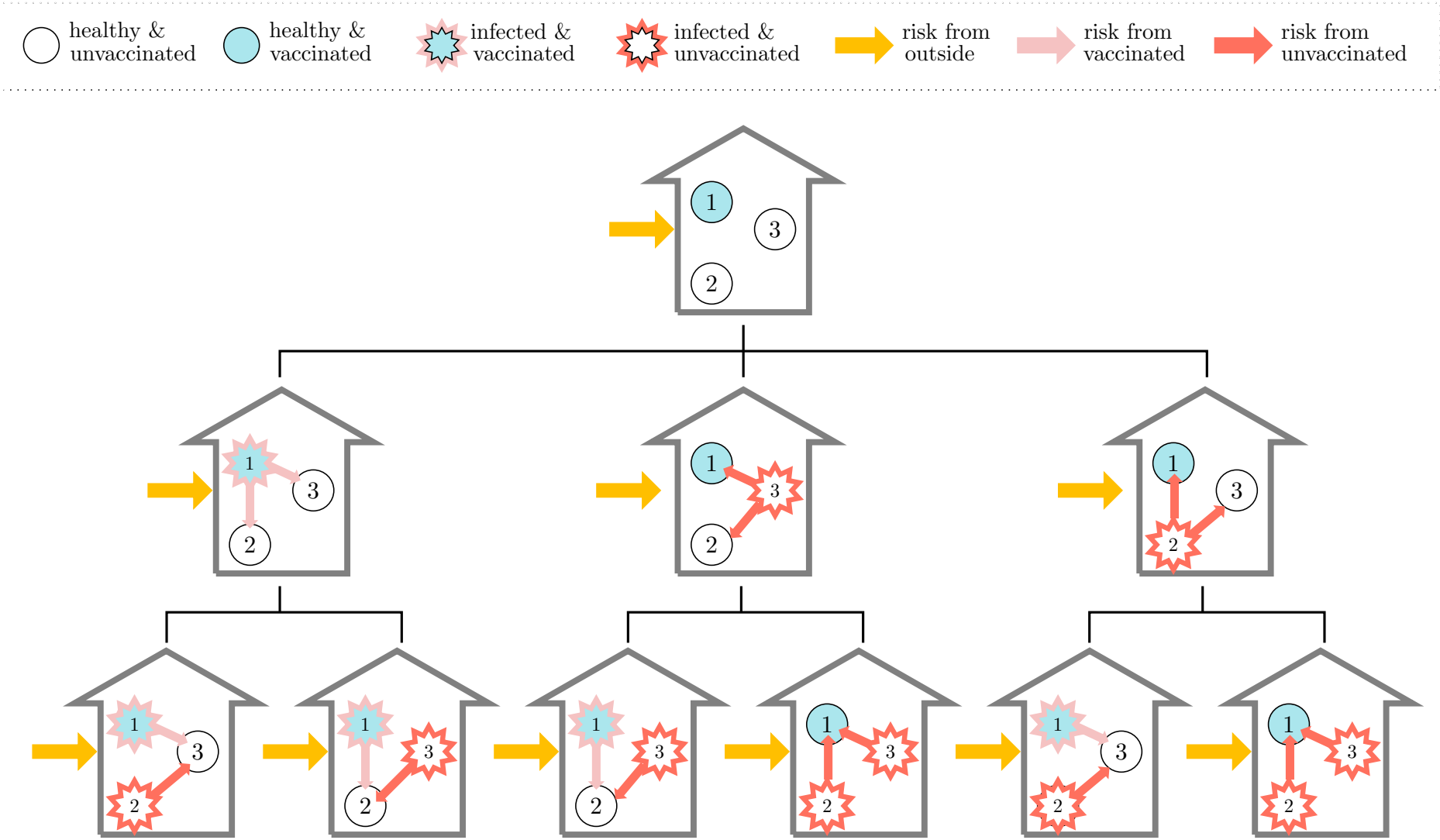
Decompose infection process regarding different orders



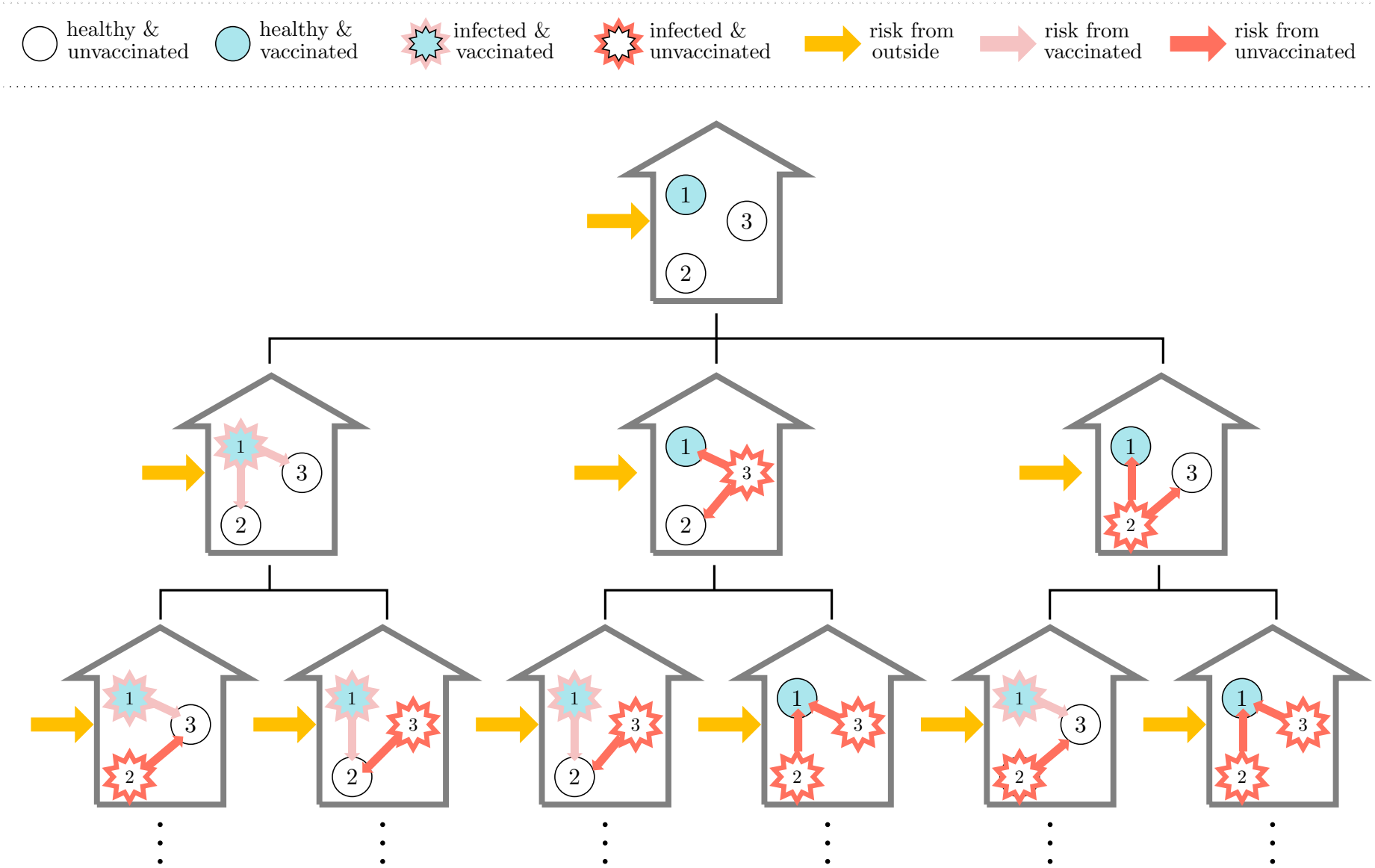
Decompose infection process regarding different orders



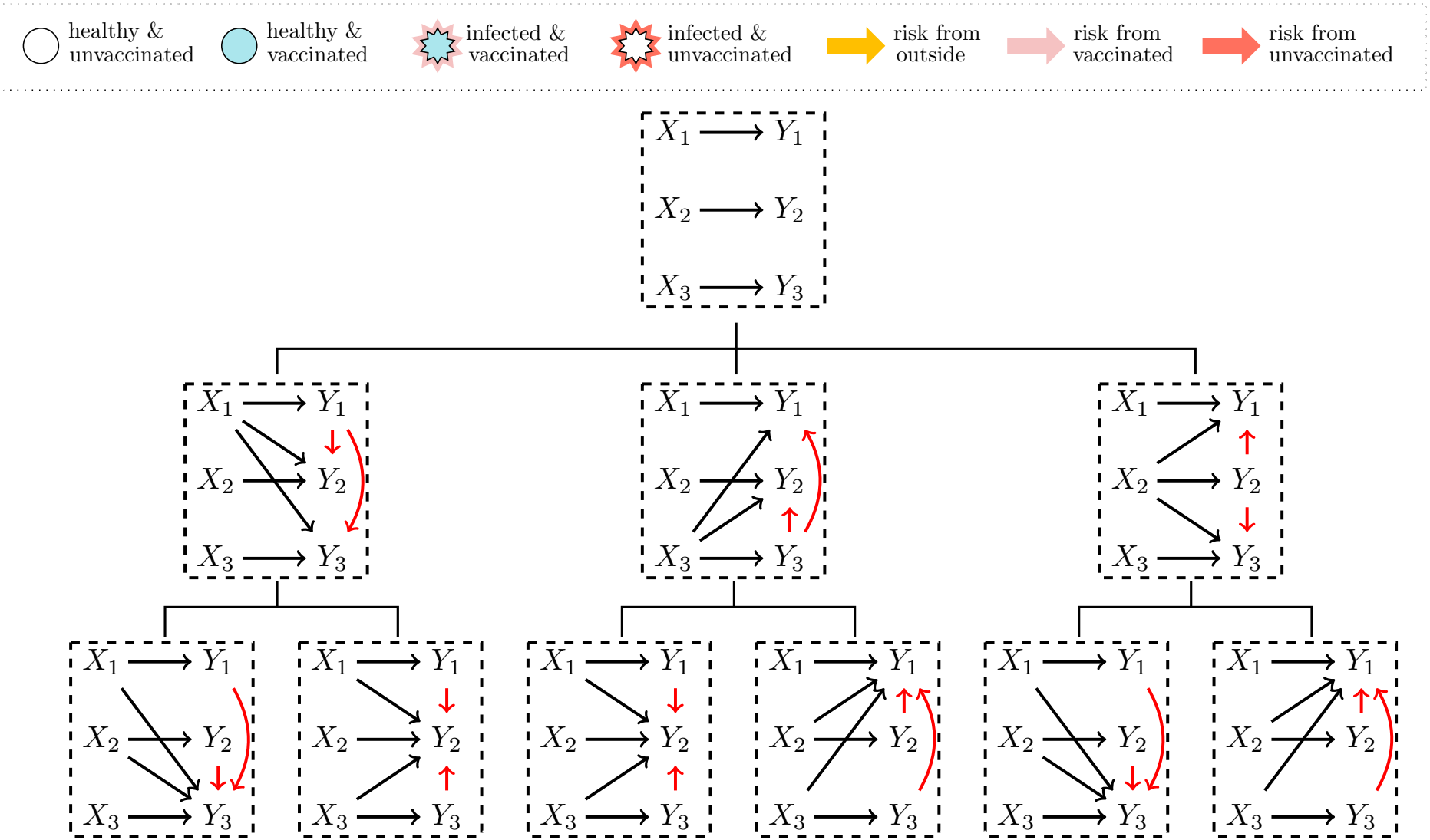
Decompose infection process regarding different orders



Decompose infection process regarding different orders



Decompose infection process regarding different orders



Notation

- Subject i's treatment: X_i
- Others' treatments: $\mathbf{X}_{(i)} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$
- Others' infection history: $\mathcal{H}_{(i)}(\mathbf{x}_{(i)}) = \{T_j(\mathbf{x}_{(i)}) : j \neq i\}$

Potential outcome

Identify $Y_i\big(t; x_i, \mathbf{x}_{(i)}, \mathcal{H}_{(i)}(\mathbf{x}_{(i)})\big)$ as the counterfactual infection outcome at time t under joint treatment $(x_i, \mathbf{x}_{(i)})$ and infection times of others $\mathcal{H}_{(i)}(\mathbf{x}_{(i)})$ under treatments $\mathbf{x}_{(i)}$,

- (i) own treatment: $X_i = x_i$
- (ii) others' treatments: $\mathbf{X}_{(i)} = \mathbf{x}_{(i)}$
- (iii) others' infection times: $\mathcal{H}_{(i)}(\mathbf{x}_{(i)})$

Identification of exposure-marginalized potential outcomes

Causal identification

Under conventional assumptions in causal inference, the potential outcome $\mathbb{E}\big[Y_i\big(t; x_i, \mathbf{x}_{(i)}, \boldsymbol{\mathcal{H}}^*_{(i)}(\mathbf{x}'_{(i)})\big)|\mathbf{L}=\mathbf{l}\big]$ can be identified as

$$\mathbb{E}\big[Y_i\big(t; x_i, \mathbf{x}_{(i)}, \boldsymbol{\mathcal{H}}^*_{(i)}(\mathbf{x}'_{(i)})\big)|\mathbf{L}=\mathbf{l}\big] = \int \mathbb{E}[Y_i(t; x_i, \mathbf{x}_{(i)}, \mathbf{h}_{(i)})|\mathbf{L}=\mathbf{l}]\,\mathrm{d}G^*_{(i)}(\mathbf{h}_{(i)}|\mathbf{x}'_{(i)}, \mathbf{l}_{(i)})$$

where

$$\mathbb{E}\big[Y_i(t; \mathbf{h}_{(i)}, \mathbf{x}) \,\big|\, \mathbf{L}=\mathbf{l}\big] = \sum_{j=0}^{n-1}\left[F_{I_i^j}(\min\{t, t^{j+1}_{(i)}\} - t^j_{(i)} \,|\, \mathbf{x}, \mathbf{h}_{(i)}, \mathbf{l}) \prod_{k=0}^{j-1}\left(1 - F_{I_i^k}(t^{k+1}_{(i)} - t^k_{(i)} \,|\, \mathbf{x}, \mathbf{h}_{(i)}, \mathbf{l})\right)\right]$$

$$\mathrm{d}G^*_{(i)}(\mathbf{h}_{(i)} \,|\, \mathbf{x}_{(i)},) = \prod_{j=1}^{n-1}\left[f_{I^{j-1}_{\varphi_i^j}}\big(t^j_{(i)} - t^{j-1}_{(i)} \,\big|\, \mathbf{x}, \mathbf{h}^i_{(\varphi_i^j)}, \mathbf{l}\big) \prod_{k=j+1}^{n-1} S_{I^{j-1}_{\varphi_i^k}}\big(t^j_{(i)} - t^{j-1}_{(i)} \,\big|\, \mathbf{x}, \mathbf{h}^j_{(\varphi_i^k)}, \mathbf{l}\big)\right]$$

$$F_{I_i^k}(s|\mathbf{x}, \mathbf{h}_{(i)}, \mathbf{l}) = 1 - \exp\left[-\int_{t^k_{(i)}}^{t^k_{(i)}+s} \frac{f_i^k(u|\mathbf{x}, \mathbf{h}_{(i)}, \mathbf{l})}{S_i^k(u|\mathbf{x}, \mathbf{h}_{(i)}, \mathbf{l})} du\right] \text{ for } k = 0, \ldots, n-1$$

Exposure-marginalized (natural) causal estimands

Exposure-marginalized (natural) causal estimands

- Susceptibility effect

$$SE_i(t, \mathbf{x}_{(i)}) = \mathbb{E}\big[Y_i\big(t; 1, \mathbf{x}_{(i)}, \boldsymbol{\mathcal{H}}^*_{(i)}(\mathbf{x}_{(i)})\big) - Y_i\big(t; 0, \mathbf{x}_{(i)}, \boldsymbol{\mathcal{H}}^*_{(i)}(\mathbf{x}_{(i)})\big)\big]$$

- Infectiousness effect

$$IE_i(t, x_i, \mathbf{x}_{(i)}) = \mathbb{E}\big[Y_i\big(t; x_i, \mathbf{1}, \boldsymbol{\mathcal{H}}_{(i)}(\mathbf{x}_{(i)})\big) - Y_i\big(t; x_i, \mathbf{0}, \boldsymbol{\mathcal{H}}_{(i)}(\mathbf{x}_{(i)})\big)\big]$$

- Contagion effect

$$CE_i(t, x_i, \mathbf{x}_{(i)}, \mathbf{x}'_{(i)}) = \mathbb{E}\big[Y_i\big(t; x_i, \mathbf{x}_{(i)}, \boldsymbol{\mathcal{H}}^*_{(i)}(\mathbf{x}_{(i)})\big) - Y_i\big(t; x_i, \mathbf{x}_{(i)}, \boldsymbol{\mathcal{H}}^*_{(i)}(\mathbf{x}'_{(i)})\big)\big]$$

- Susceptibility effect $\rightarrow$ shows if the vaccine protects treated individual
- Infectiousness effect $\rightarrow$ shows if the vaccine decreases transmission ability
- Contagion effect $\rightarrow$ shows if the disease is contagious

Traditional causal estimands in cluster studies

- Direct effect:

$$DE(t) = \mathbb{E}[Y_i(t)|X_i = 1] - \mathbb{E}[Y_i(t)|X_i = 0]$$

- Indirect effect:

$$IDE(t) = \sum_{|\mathbf{x}_{(i)}|=\frac{n}{2}} \mathbb{E}[Y_i(t)|X_i = 0, \mathbf{X}_{(i)} = \mathbf{x}_{(i)}]p(\mathbf{x}_{(i)}) - \sum_{|\mathbf{x}_{(i)}|=0} \mathbb{E}[Y_i(t)|X_i = 0, \mathbf{X}_{(i)} = \mathbf{x}_{(i)}]p(\mathbf{x}_{(i)})$$

[1] Longini et al. Statistical inference for infectious diseases: risk-specific household and community transmission parameters. American Journal of Epidemiology, 128(4):845–859, 1988.
[2] Halloran et al. Direct and indirect effects in vaccine efficacy and effectiveness. American Journal of Epidemiology, 133(4):323–331, 1991.
[3] Halloran et al. Exposure efficacy and change in contact rates in evaluating prophylactic HIV vaccines in the field. Statistics in Medicine, 13(4):357–377, 1994.

Simulation: true values of causal estimands

Cluster	Treatment	Causal estimands				
		CE	SE	IE	$DE(t)$	$IDE(t)$
2	Obs.	0.004	-0.013	-0.037	-0.012	-
	Bernoulli	0.004	-0.013	-0.037	-0.012	-0.037
	Block	-	-	-	0.024	-
	Cluster	-	-	-	-0.049	-
4	Obs.	0.026	-0.013	-0.084	-0.011	-
	Bernoulli	0.026	-0.013	-0.084	-0.011	-0.063
	Block	-	-	-	0.018	-
	Cluster.	-	-	-	-0.098	-
8	Obs.	0.066	-0.013	-0.129	-0.010	-
	Bernoulli	0.066	-0.013	-0.129	-0.010	-0.094
	Block	-	-	-	0.010	-
	Cluster	-	-	-	-0.151	-

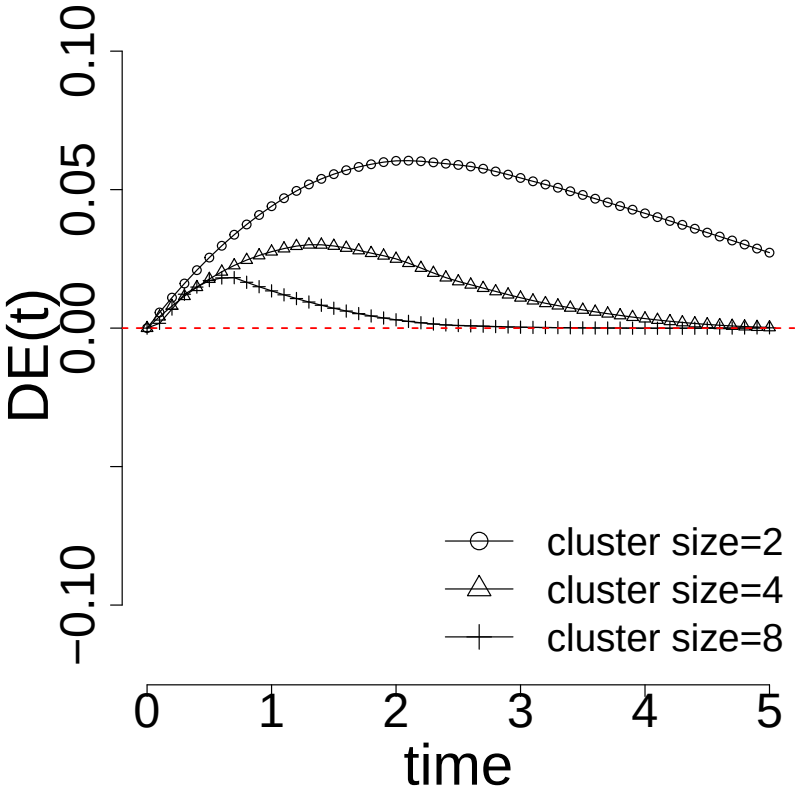
Simulation under $e^{\beta_1} = 0.9$, $e^{\beta_2} = 0.1$, $\alpha(t) = 0.3$, $\gamma(t) = 3$ and $e^{\theta_1} = e^{\theta_2} = 0.9$. Clusters of 2, 4, and 8 are observed at 0.4, 0.3 and 0.2 year.

Simulation: true values of causal estimands

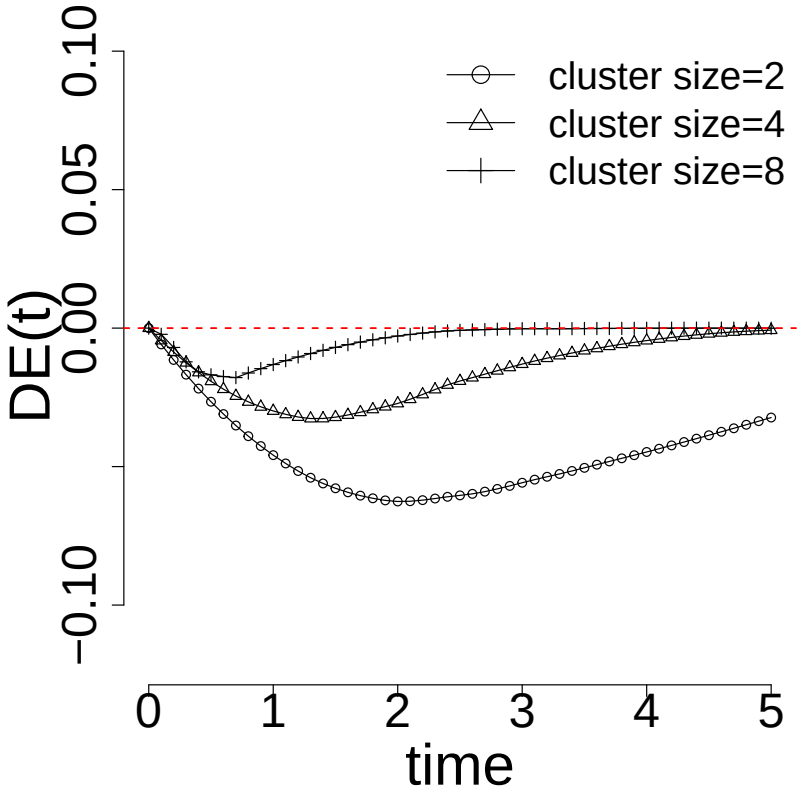
Cluster	Treatment	Causal estimands				
		$CE(t)$	$SE(t)$	$IE(t)$	$DE(t)$	$IDE(t)$
2	Obs.	0.004	-0.013	-0.037	-0.012	-
	Bernoulli	0.004	-0.013	-0.037	-0.012	-0.037
	Block	-	-	-	0.024	-
	Cluster	-	-	-	-0.049	-
4	Obs.	0.026	-0.013	-0.084	-0.011	-
	Bernoulli	0.026	-0.013	-0.084	-0.011	-0.063
	Block	-	-	-	0.018	-
	Cluster.	-	-	-	-0.098	-
8	Obs.	0.066	-0.013	-0.129	-0.010	-
	Bernoulli	0.066	-0.013	-0.129	-0.010	-0.094
	Block	-	-	-	0.010	-
	Cluster	-	-	-	-0.151	-

Simulation under $e^{\beta_1} = 0.9$, $e^{\beta_2} = 0.1$, $\alpha(t) = 0.3$, $\gamma(t) = 3$ and $e^{\theta_1} = e^{\theta_2} = 0.9$. Clusters of 2, 4, and 8 are observed at 0.4, 0.3 and 0.2 year.

Simulation: correlated treatment assignment under null direct protection



Negatively correlated



Positively correlated

Simulation under $e^{\beta_1} = 1$, $e^{\beta_2} = 0.1$, $\alpha(t) = 0.3$, and $\gamma(t) = 3$.

Estimation

Treatment	$\hat{CE}$	$\hat{SE}$	$\hat{IE}$	$\hat{DE}(t)$	$\hat{IDE}(t)$
Obs.	0.265	-0.075	-0.047	-0.105	-
	(0.189,0.345)	(-0.122,-0.033)	(-0.088,-0.013)	(-0.150,-0.036)	-
Bernoulli	0.257	-0.077	-0.046	-0.102	-0.147
	(0.182,0.335)	(-0.124,-0.035)	(-0.088,-0.013)	(-0.151,-0.052)	(-0.231,-0.060)
Block	-	-	-	-0.065	-
	-	-	-	(-0.094,-0.038)	-
Cluster	-	-	-	-0.448	-
	-	-	-	(-0.566,-0.328)	-

Estimation with 100 clusters of size 10 at t=0.5, when $e^{\beta_1} = 0.5$, $e^{\beta_2} = 0.5$, $\alpha(t) = 0.3$, $\gamma(t) = 3$ and $e^{\theta_1} = e^{\theta_2} = 0.9$. True values are $SE(t = 0.5) = 0.265$, $IE(0.5) = -0.075$, $CE(t = 0.5) = -0.048$ across various treatment assignments, and $IDE(t) = -0.151$ under Bernoulli randomization. True values of $DE(0.5)$ are -0.107, -0.107, -0.067, and -0.458, respectively for Obs, Bernoulli, Block, and Cluster randomization.

Estimation

Treatment	$\hat{CE}$	$\hat{SE}$	$\hat{IE}$	$\hat{DE}(t)$	$\hat{IDE}(t)$
Obs.	0.265	-0.075	-0.047	-0.105	-
	(0.189,0.345)	(-0.122,-0.033)	(-0.088,-0.013)	(-0.150,-0.036)	-
Bernoulli	0.257	-0.077	-0.046	-0.102	-0.147
	(0.182,0.335)	(-0.124,-0.035)	(-0.088,-0.013)	(-0.151,-0.052)	(-0.231,-0.060)
Block	0.255	-0.076	-0.047	-0.065	-
	(0.179,0.332)	(-0.122,-0.035)	(-0.091,-0.012)	(-0.094,-0.038)	-
Cluster	0.253	-0.079	-0.050	-0.448	-
	(0.150,0.359)	(-0.146,-0.024)	(-0.101,-0.011)	(-0.566,-0.328)	-

Estimation with 100 clusters of size 10 at t=0.5, when $e^{\beta_1} = 0.5$, $e^{\beta_2} = 0.5$, $\alpha(t) = 0.3$, $\gamma(t) = 3$ and $e^{\theta_1} = e^{\theta_2} = 0.9$. True values are $SE(t = 0.5) = 0.265$, $IE(0.5) = -0.075$, $CE(t = 0.5) = -0.048$ across various treatment assignments, and $IDE(t) = -0.151$ under Bernoulli randomization. True values of $DE(0.5)$ are -0.107, -0.107, -0.067, and -0.458, respectively for Obs, Bernoulli, Block, and Cluster randomization.

Summary

- We articulate the causal structure between individuals' treatments and outcomes of infectious disease in a inter-connected cluster.
- We illustrate the identification strategy for infectious disease potential outcomes non-parametrically
  - Do not depend on certain study design or randomization strategy
  - Apply to various transmission dynamics, cluster size and observational time
  - Incorporate individual- and cluster-level covariates
- A class of novel and biologically meaningful causal estimands for the susceptibility, infectiousness and contagion effect of vaccines are proposed
- We provide comprehensive comparisions to popular estimands in contemporary epidemiology

Other relevant work and future direction

- We further apply a generalized Cox-type transmission hazard model to facilitate the inference of causal estimands parametrically or semi-parametrically.
- We promote hazard ratio as alternative causal estimands for the susceptibility and infectiousness effect, and compared them to existing estimands for vaccine efficacy.
- Extend current research on causal identification for contagious outcomes to more realistic scenarios, for example, relaxing requirement on accurate infection times, accommodating incomplete knowledge of transmission network, allowing recovering and re-infection of outcomes.

References

Xiaoxuan Cai, Wen Wei Loh, Forrest W. Crawford. (2021) Identification of Causal intervention effects under contagion. Journal of Causal Inference, 9, 9-38. (Winner of best paper award, ASA Section on Statistics in Epidemiology)

Xiaoxuan Cai, Eben Kenah, Forrest W. Crawford. (2021) Causal identification of infectious disease intervention effects in a clustered population. arXiv:2105.03493

Acknowledgement

The work was supported by National Institutes of Health (NIH) grant NICHD/BD2K 1DP2HD09179 and R01AL116770.

We thank Peter Aronow, Zehang Li, Olga Morozova, Virginia Pitzer, Elizabeth Ogburn, and Li Zeng for their great help and suggestions.

cai.1083@osu.edu
https://xiaoxuan-cai.github.io/
(References and slides are available on my personal website.)

Thank you!