Causal inference for infectious disease intervention under contagion

Xiaoxuan Cai

joint work with Wen Wei Loh, Eben Kenah, Forrest W. Crawford

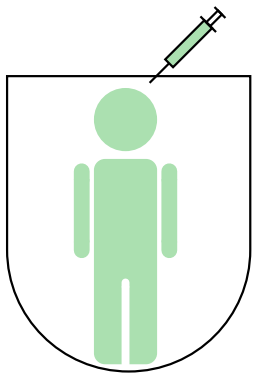
Department of Biostatistics, School of Public Health
Columbia University

October 2, 2020 WSDS

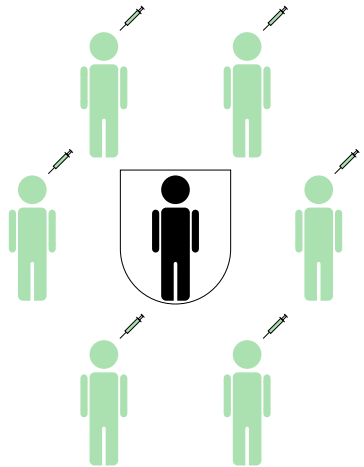
Infectious disease and vaccination

Distinct mechanisms of infectious disease interventions/vaccinations,

- Direct protection for the treated individuals:
  - direct effect, vaccine efficacy, susceptibility effect...
- Indirect protection for the surrounding individuals:
  - indirect effect, herd immunity...
- Vaccines for Polio, Influenza, HIV/AIDS, Malaria and etc.



Direct protection



Indirect protection

Why infectious disease is difficult to study?

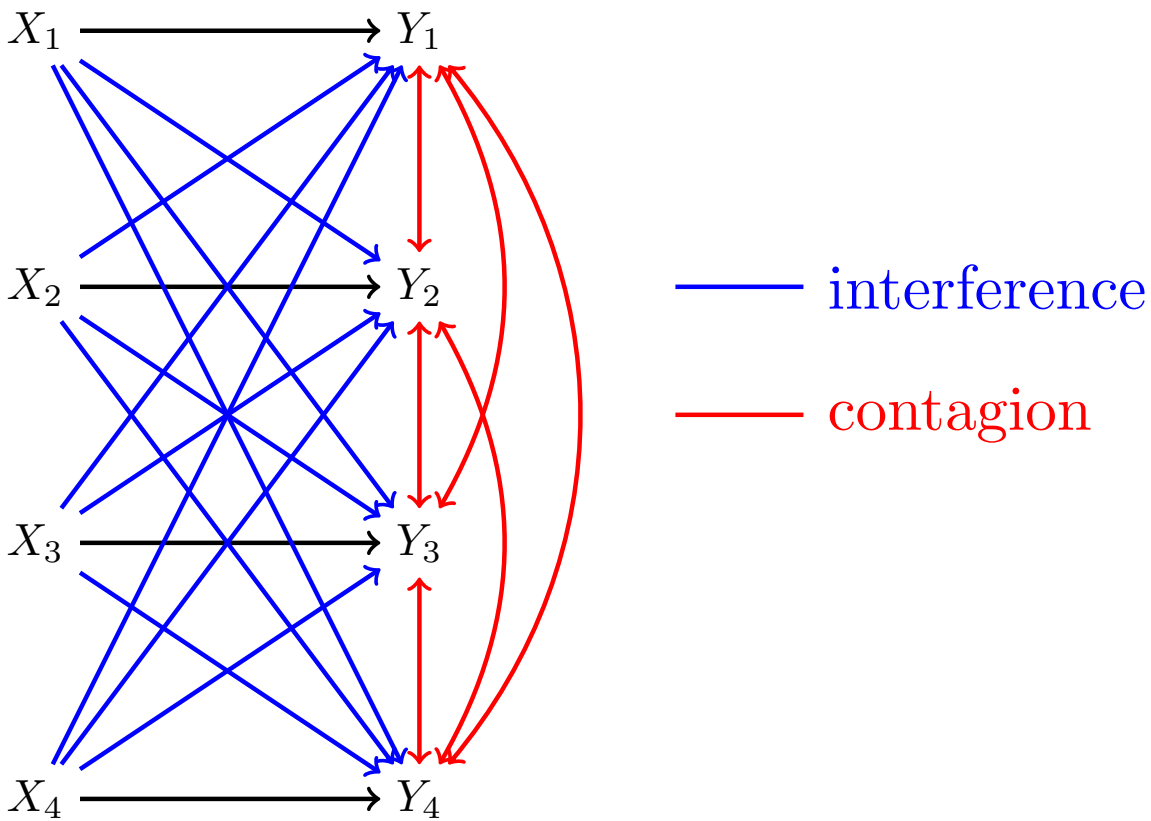
Research on transmission of infectious disease has some unique features and challenges.

- The infection outcome of one individual also depends on others' treatments, conditional on other individuals being infected. – Interference
- The outcome of interest (infection) is transmissible, so outcomes are not independent from each other. – Contagion
- The infection times of others compose an important factor for the infection outcome – exposure to infection
- Individuals' interaction along the transmission process reveals essential information about transmission mechanism.

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

Challenges for causal identification

Consider a interconnected four individuals with treatment (X_1, X_2, X_3, X_4) and infection outcome (Y_1, Y_2, Y_3, Y_4).



- The graph is not an acyclic directed graph (DAG).

How to solve the problem?

Can randomization solve the problem?

- Even under randomization, direct comparisons of treated and untreated individuals may not be valid due to differential "exposure to infection".
- For example, if vaccinated individuals get infected later in general, then later infected, vaccinated subjects face higher exposure to infection, comparing to unvaccinated, earlier infected individuals.
  $\rightarrow$ not a fair comparison !

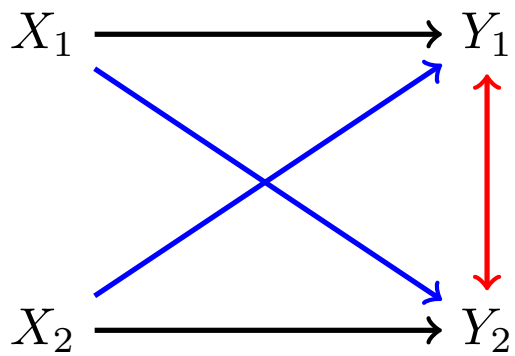
Propose new methodology to evaluate interventions effects for contagious outcomes

We will provide new methods that

- 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
- Yield biologically meaningful causal estimands for direct and indirect protection provided by interventions
- Allow flexible statistical inferential framework, ranging from parametric, semi-parametric to non-parametric estimation

Symmetric partnership models

Partnership models have been widely understood as a useful framework to clarify causal relationship in epidemiology, and lay the foundation for more complex settings.
Consider two individuals with treatment X_1 and X_2 and infection outcome Y_1 and Y_2.



- The graph is not an acyclic directed graph (DAG).

Unique challenges for causal identification under contagion

Problem: Differentiating exposure to infection after randomization

Solution:

Add a component of "exposure to infection" (other's infection times) into the counterfactual outcome definition for a fair comparison.

Problem: Bidirectional arrow in the causal diagram

Solution:

Transform the cyclic diagram into traditional DAG by separating the transmission process into exclusive possibilities.

Notation

For the symmetric partnership models, consider individual 1 and 2 and let,

- Treatment assignment: $\mathrm{X} = (X_1, X_2)$
- Infection time: T_i for $i = 1, 2$
- Infection outcome: $Y_i(t) \equiv \mathbb{1}\{t \geq T_i\}$ for $i = 1, 2$
- Isolated infection time: I_i^0 for $i = 1, 2$
- Extra infection time after partner's infection: $I_i^1 = T_i - I_j^0$ for $i \neq j$
- Counterfactual infection outcome $Y_i(t; I_j^0 = s, \mathrm{X} = \mathrm{x})$ for $j \neq i$ and $i = 1, 2$, when we fix $I_j^0 = s$ and $\mathrm{X} = \mathrm{x}$.

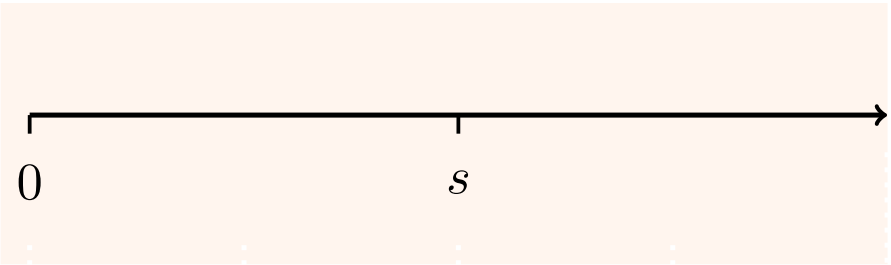
Goal:

Identify $Y_i(t; s, \mathrm{x})$ under joint intervention (s, x) for $i = 1, 2$

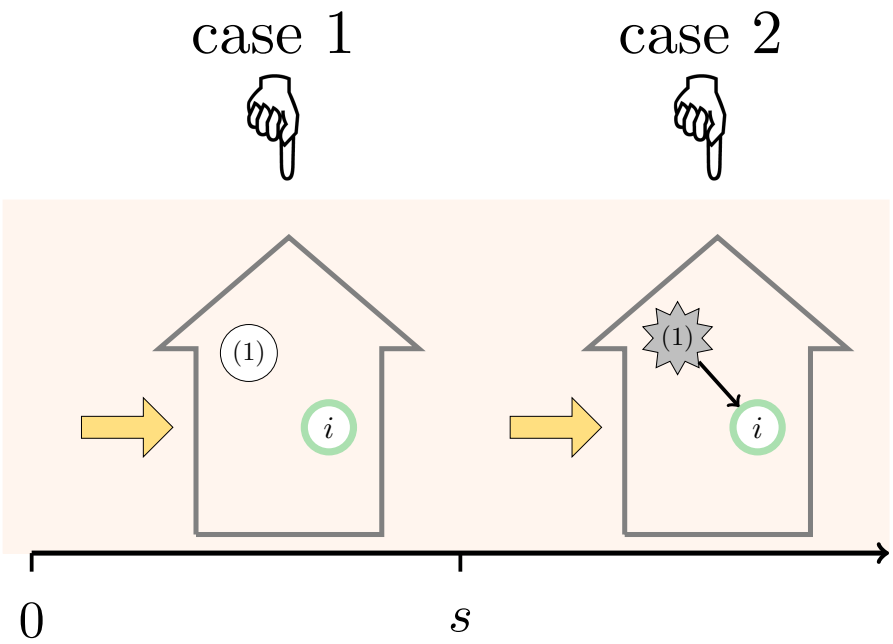
Clusters of 2 individuals

case 1

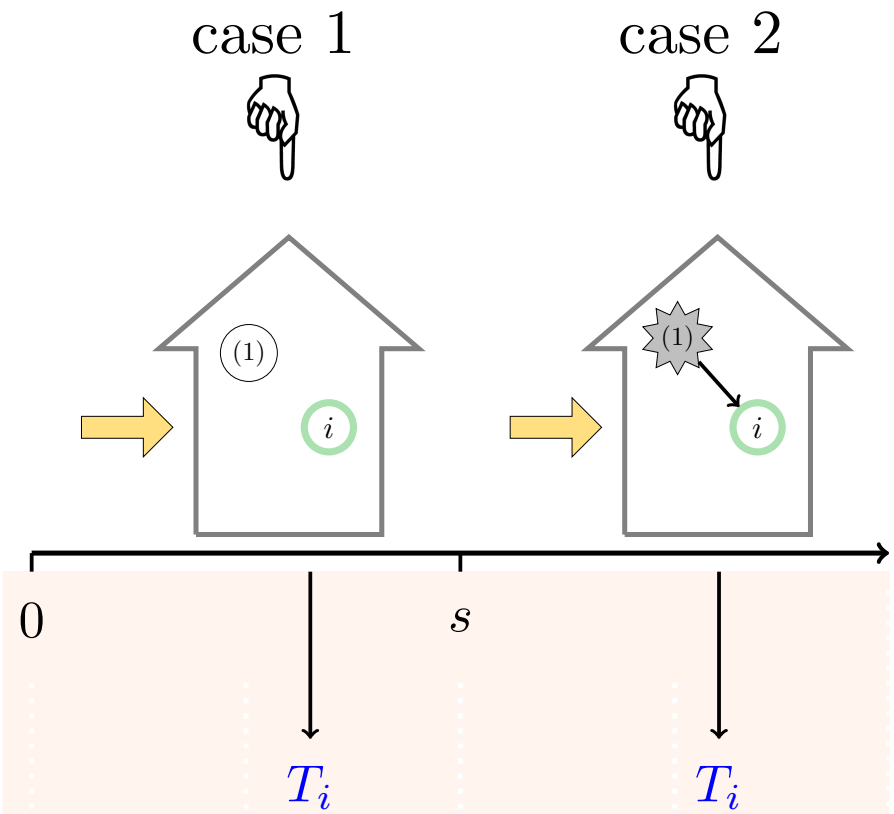
case 2



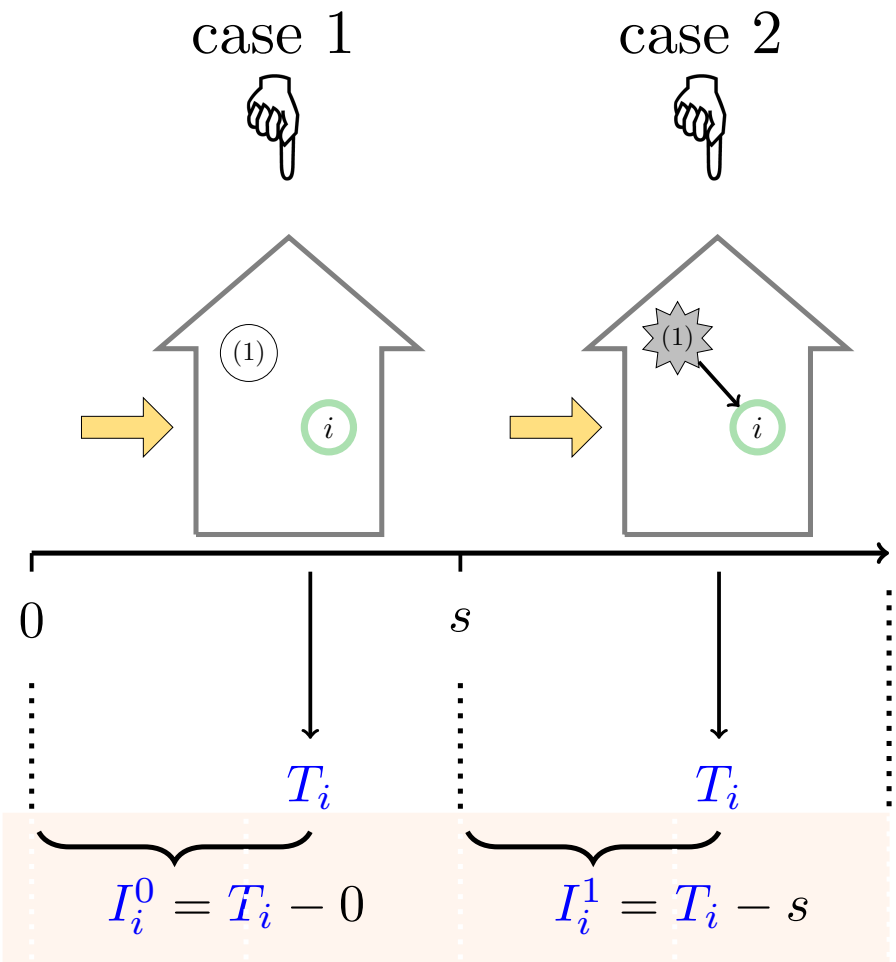
Clusters of 2 individuals



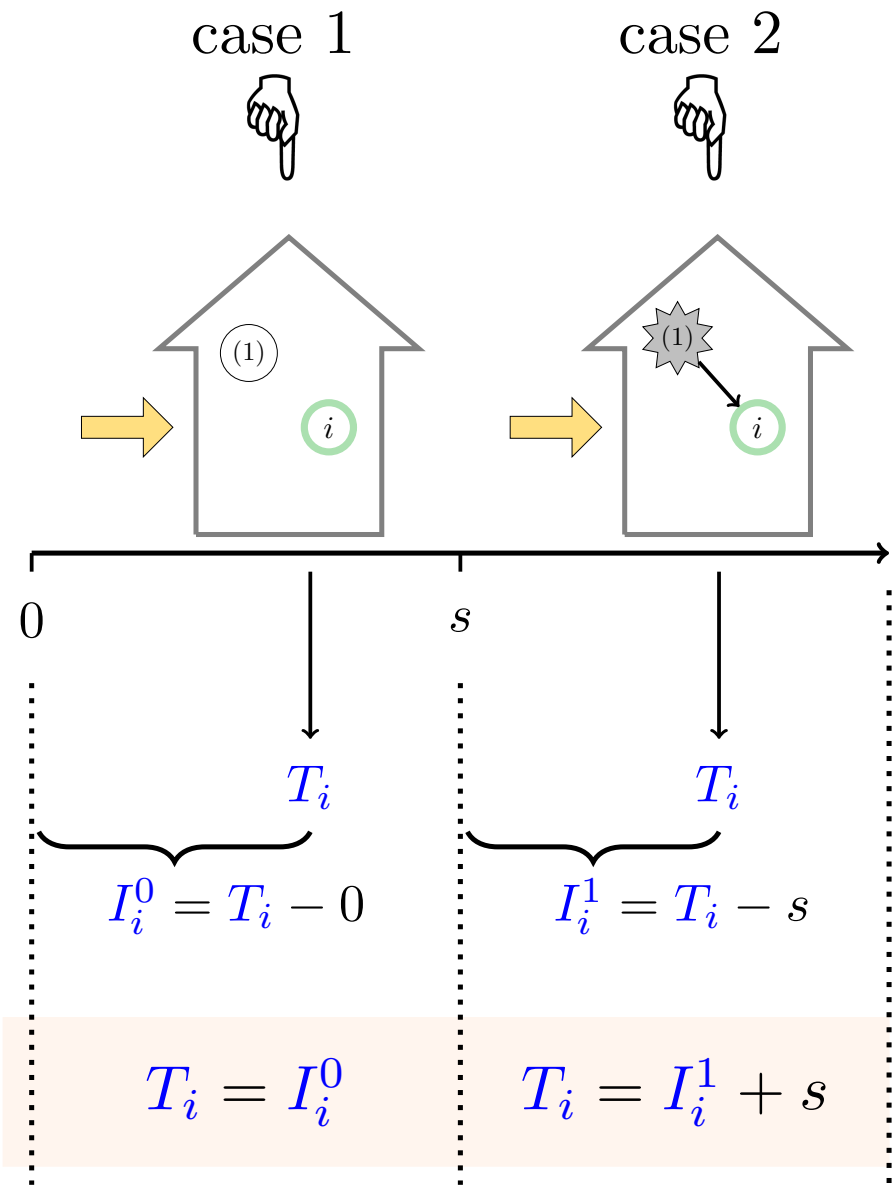
Clusters of 2 individuals



Clusters of 2 individuals



Clusters of 2 individuals



Main Result

Use subject i as a focal subject

$$T_2(\mathrm{x}) \quad = \quad \begin{cases} I_i^0(x_i) & \text{if } I_i^0(x_i) < I_j^0(x_j) \\ s + I_i^1(t_{(1)}; \mathrm{x}) & \text{otherwise} \end{cases}$$

Theorem

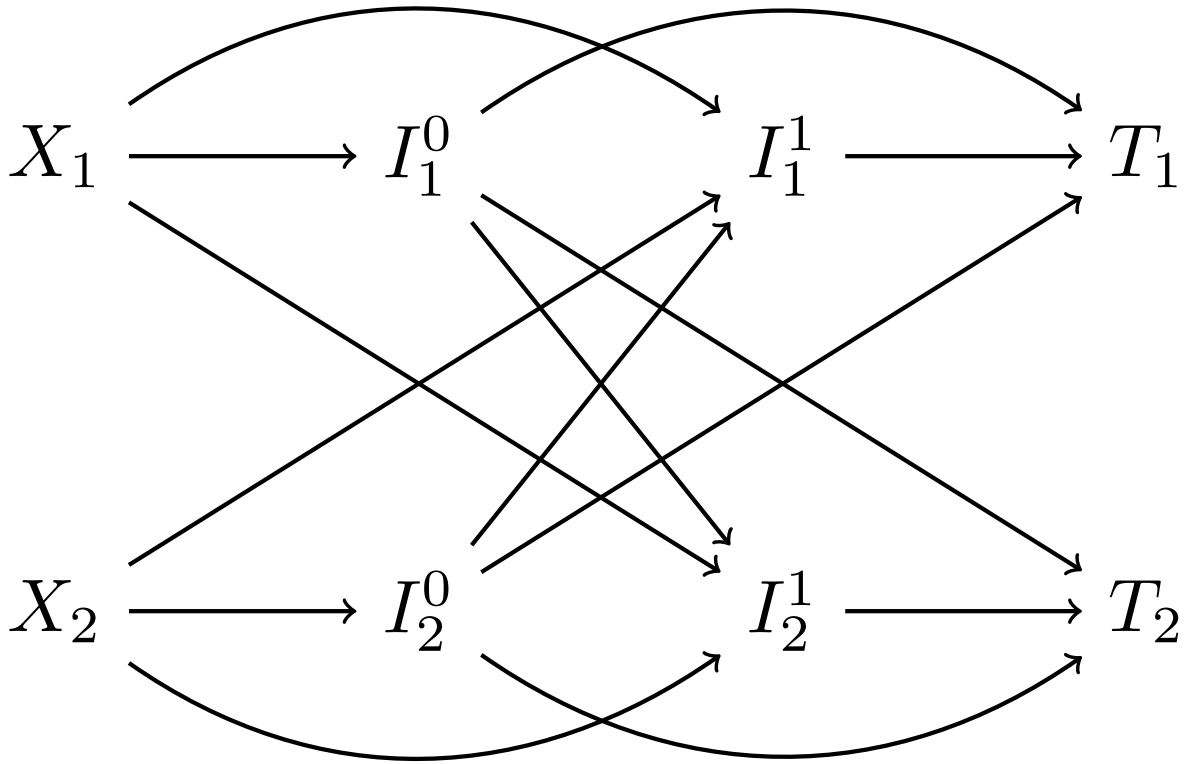
The average risk of infection by time t for the focal individual i, which is $\mathbb{E}[Y_i(t; s, \mathrm{x})]$, is identified as:

$$\mathbb{E}[Y_2(t; s, \mathrm{x})] = 1 \cdot p_2(s|\mathrm{x}) + \mathbb{E}[Y_2(t)|\, T_2 \geq s,\, T_1 = s, \mathrm{X} = \mathrm{x}] \cdot [1 - p_2(s|\mathrm{x})]$$

where $p_2(s|\mathrm{x}) = 1 - \exp[-\int_0^s \frac{\Pr(T_2 = u, T_1 > u|\mathrm{X}=\mathrm{x})}{\Pr(T_2 > u, T_1 > u|\mathrm{X}=\mathrm{x})} du]$

Causal Diagram

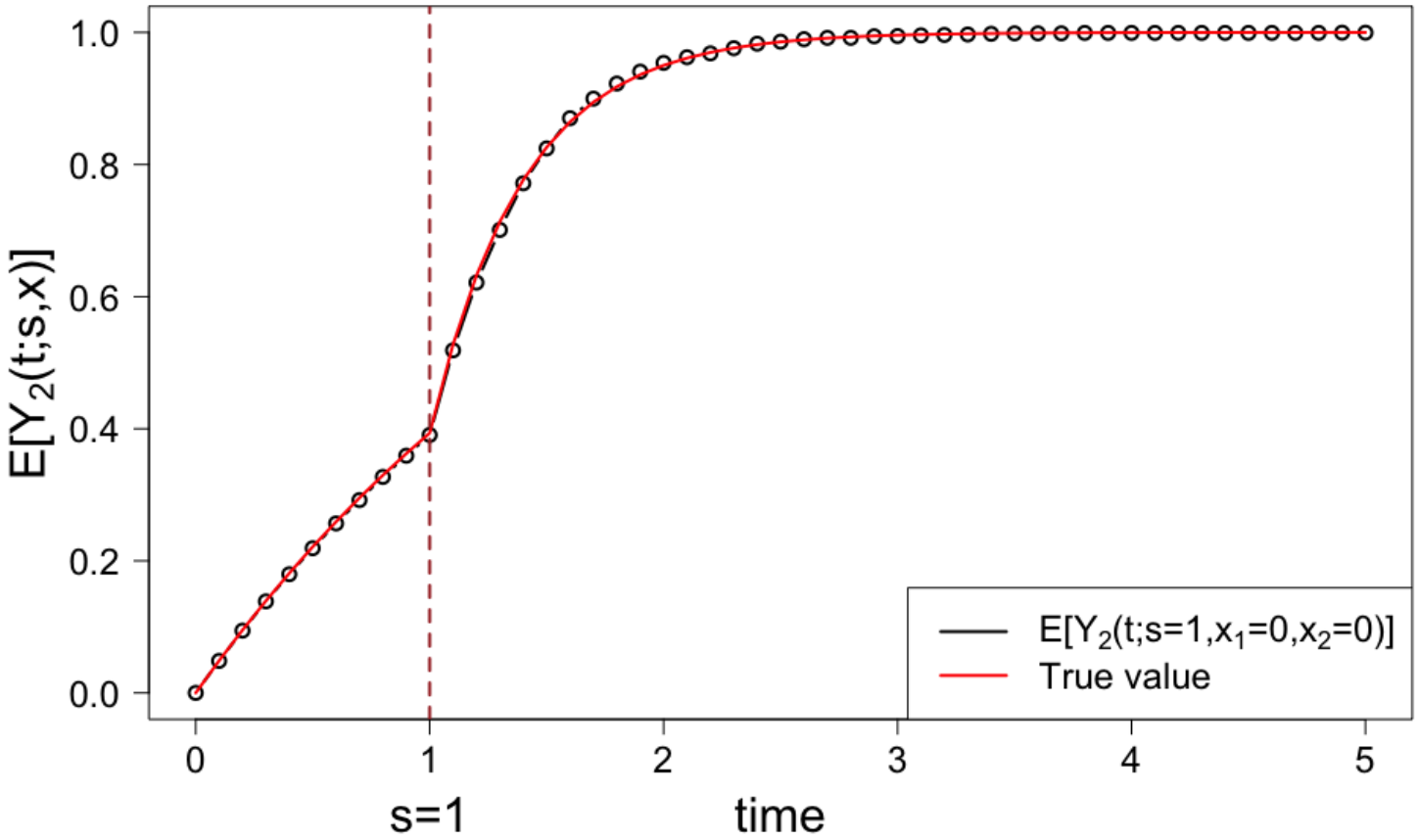
After we split T_i as W_i and Z_i, we have a acyclic causal diagram.



Note: Covariates L are omitted for simplicity of the representation.

Simulation

We simulate N=100,000 partnerships with exogenous hazard $\alpha(t) = 0.5$ and within-pair hazard $\gamma(t) = 2$. Vaccinations decrease risks by 50%, which is $e^{\beta_1} = e^{\beta_2} = 0.5$. We choose $s = 1$ and $s' = 2$.



Controlled Causal estimands

Controlled causal estimands

- Susceptibility effect ($s > 0$)

$$SE(t, s, x_1) = \mathbb{E}[Y_2(t; s, x_1, 1) - Y_2(t; s, x_1, 0)]$$

- Infectiousness effect ($s > 0$)

$$IE(t, s, x_2) = \mathbb{E}[Y_2(t; s, 1, x_2) - Y_2(t; s, 0, x_2)]$$

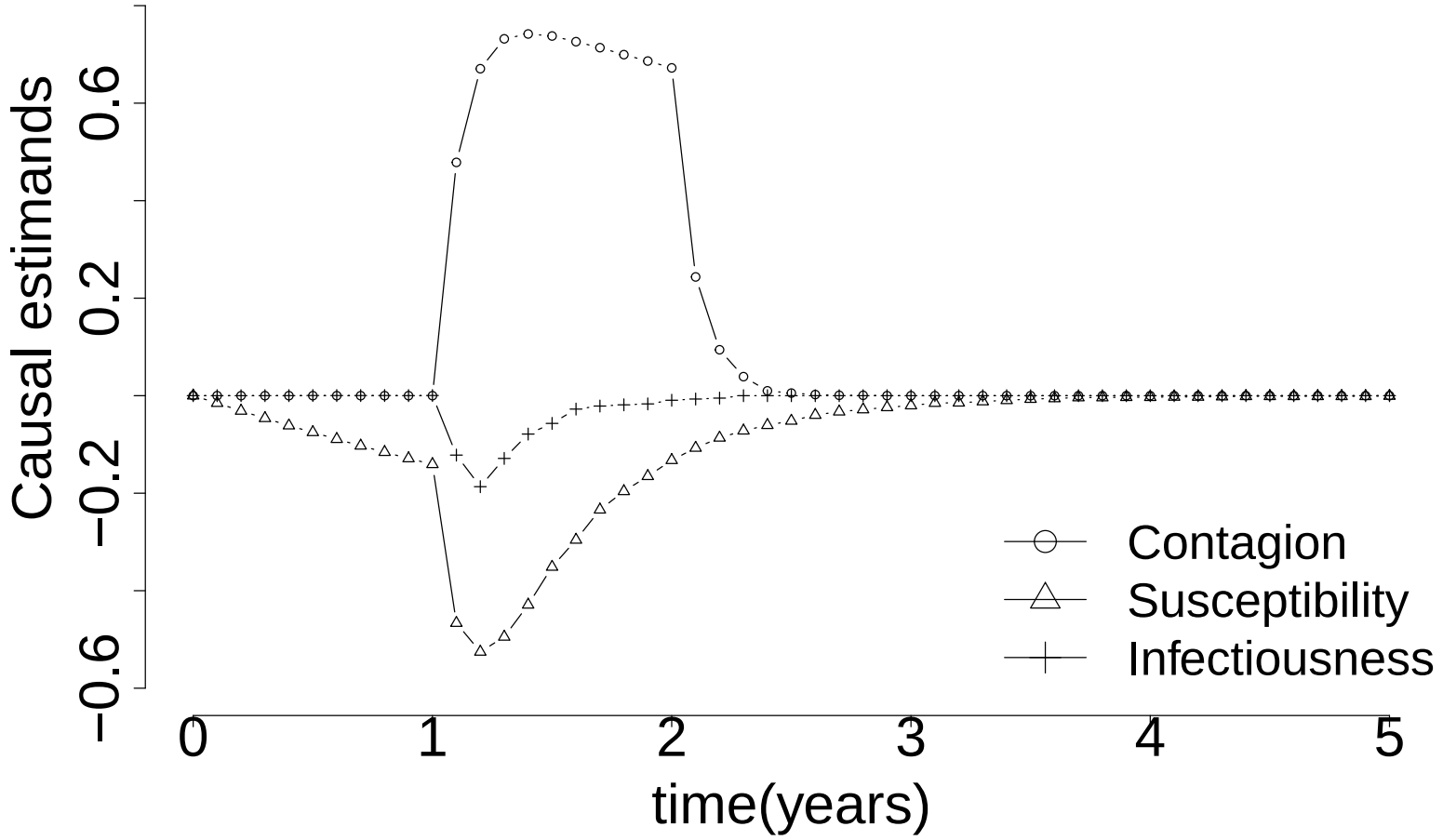
- Contagion effect ($s \neq s'$ and $\mathrm{X} = (0, 0)$)

$$CE(t, s, s') = \mathbb{E}[Y_i(t; s', 0, 0) - Y_i(t; s, 0, 0)]$$

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

Simulation: Estimates of controlled causal estimands

We simulate N=100,000 partnerships with exogenous hazard $\alpha(t) = 0.5$ and within-pair hazard $\gamma(t) = 2$. Vaccinations decrease risks by 50%, which is $e^{\beta_1} = e^{\beta_2} = 0.5$. We choose $s = 1$ and $s' = 2$.



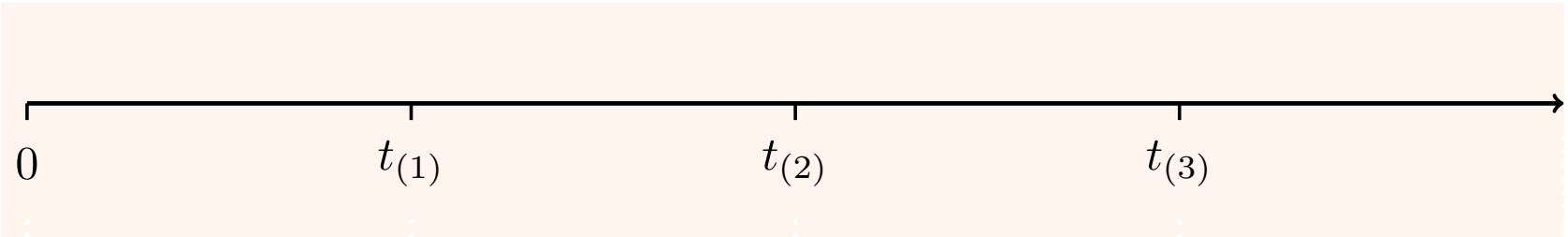
Simulation: Estimations of natural causal estimands

Simulation	Treatment	$CE(t,0,0)$	$SE(t,0)$	$IE(t,0)$	DE(t)	IDE(t)
Constant hazards	Obs.	0.12	-0.14	-0.19	-0.16	-0.20
	Bernoulli	0.12	-0.14	-0.19	-0.16	-0.20
	Block	-	-	-	0.06	-
	Cluster	-	-	-	-0.39	-
Constant hazards without contagion	Obs.	0.00	-0.18	0.00	-0.17	0.00
	Bernoulli	0.00	-0.18	0.00	-0.18	0.00
	Block	-	-	-	-0.18	-
	Cluster	-	-	-	-0.18	-
Time-varying hazards	Obs.	0.12	-0.14	-0.20	-0.22	-0.21
	Bernoulli	0.12	-0.14	-0.20	-0.21	-0.22
	Block	-	-	-	0.08	-
	Cluster	-	-	-	-0.50	-
Time-varying hazards without contagion	Obs.	0.00	-0.28	0.00	-0.28	0.00
	Bernoulli	0.00	-0.28	0.00	-0.28	0.00
	Block	-	-	-	-0.28	-
	Cluster	-	-	-	-0.28	-

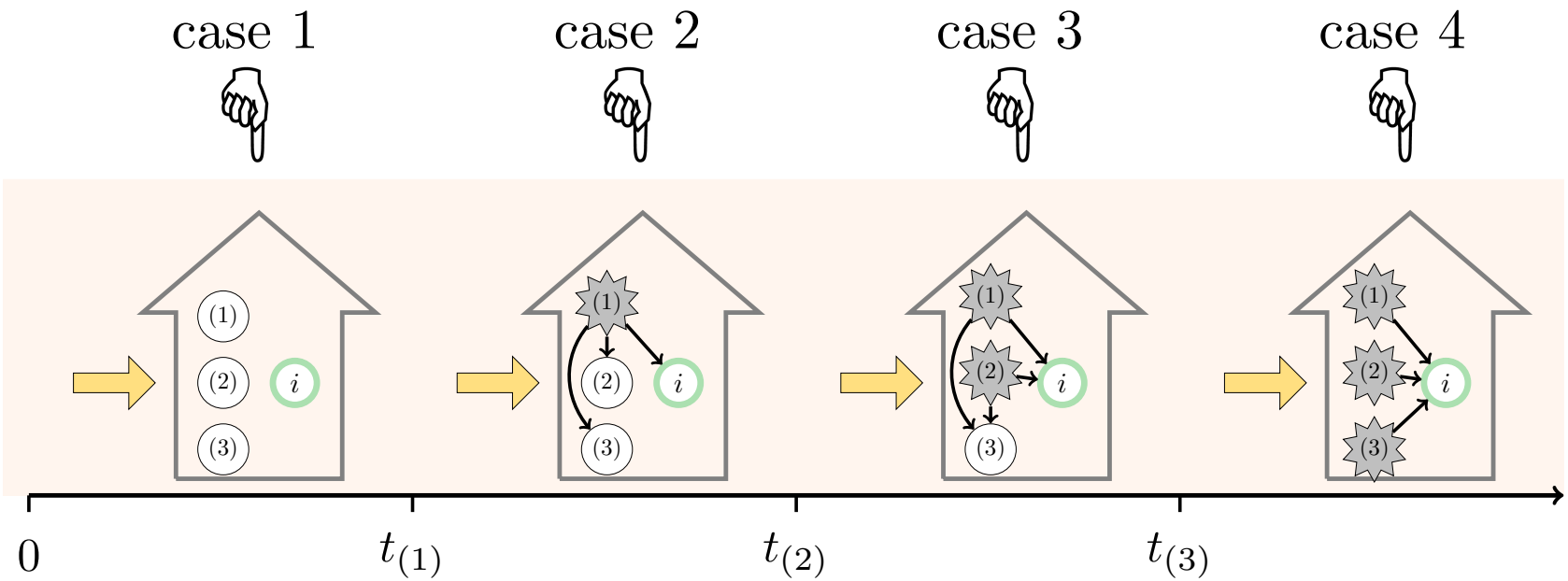
- Direct effect: $DE(t) = \mathbb{E}[Y_i(t)|X_i = 1] - \mathbb{E}[Y_i(t)|X_i = 0]$
- Indirect effect: $IDE(t) = \mathbb{E}[Y_i(t)|X_j = 1] - \mathbb{E}[Y_i(t)|X_j = 0]$

Clusters of 4 individuals

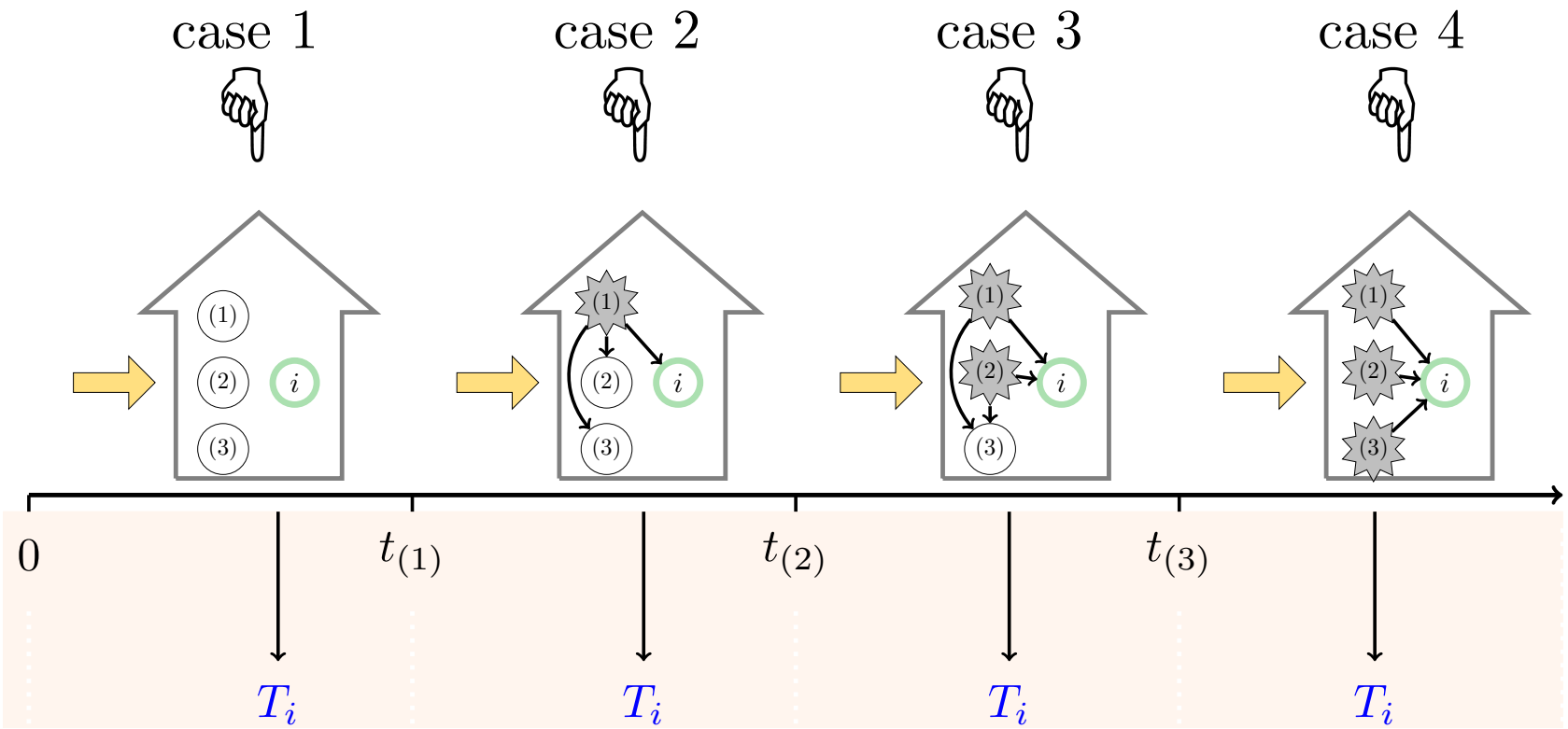
case 1 case 2 case 3 case 4



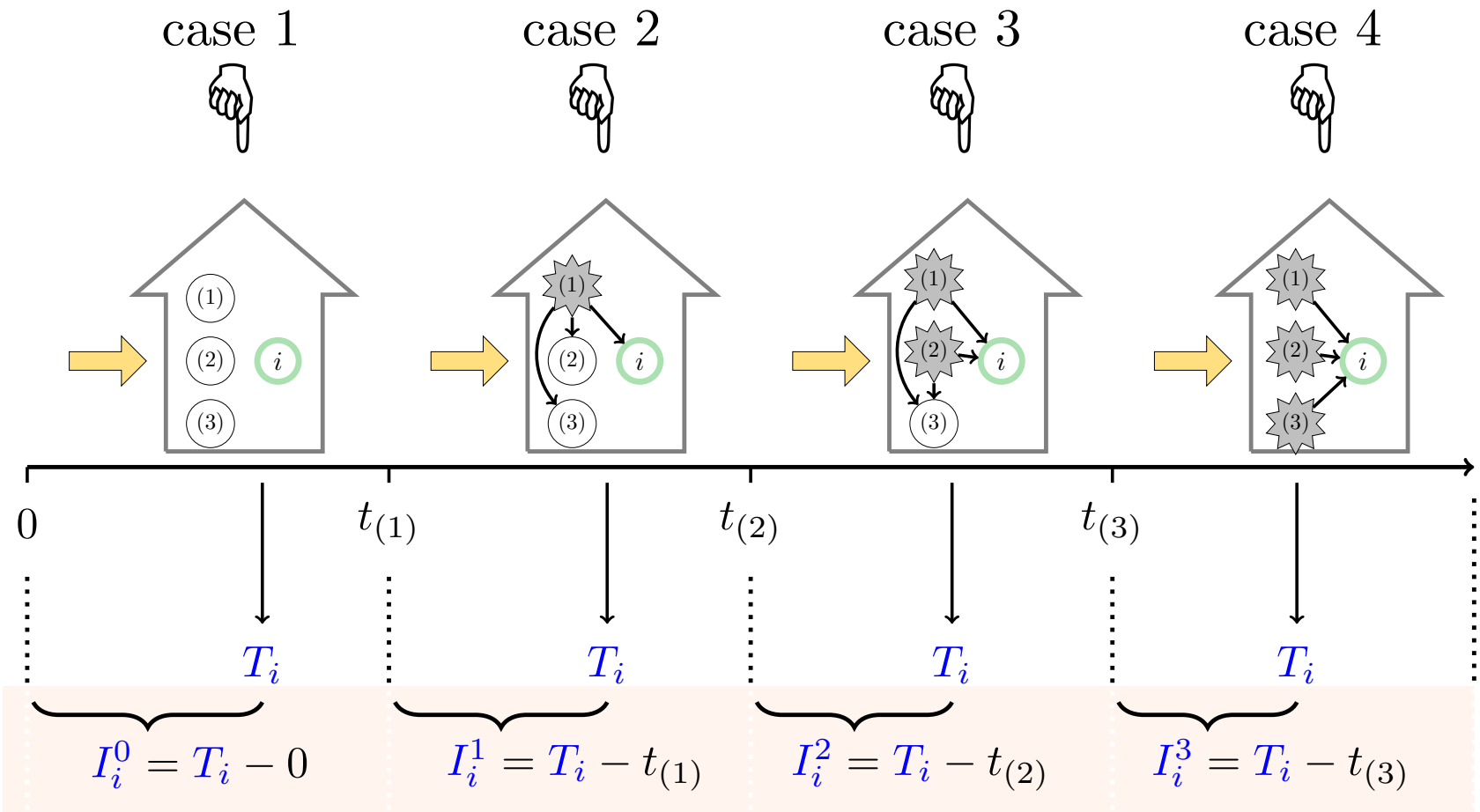
Clusters of 4 individuals



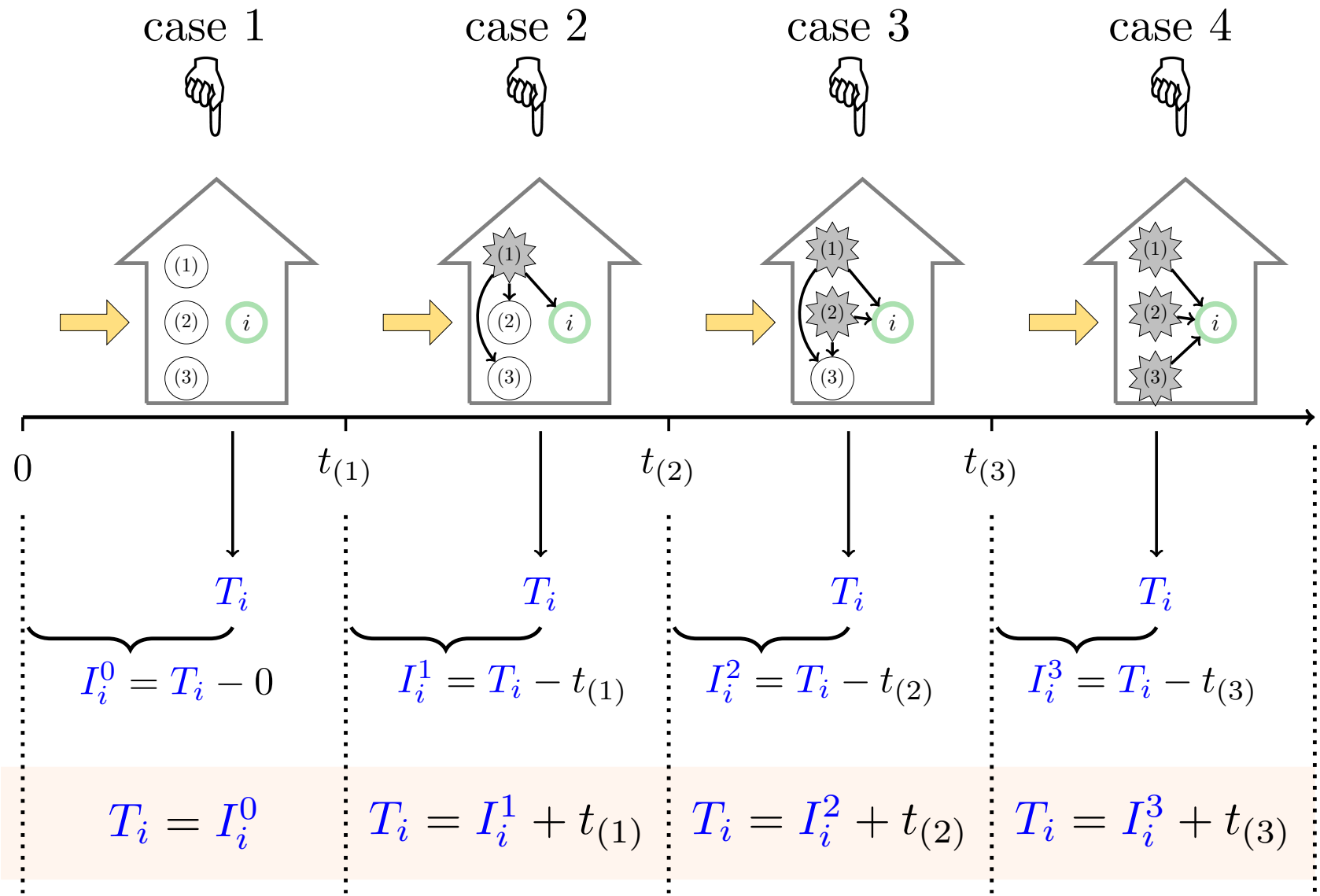
Clusters of 4 individuals



Clusters of 4 individuals



Clusters of 4 individuals



Main Result: Exposure-controlled potential outcome

Theorem: Identification of exposure-controlled potential outcomes

Under conventional assumptions in causal inference, the potential outcome under a deterministic infection times of others $\mathsf{h}_{(i)}$ and treatment x is

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

where $F_{I_i^j}(s \,|\, \mathsf{x},\mathsf{h}_{(i)},\mathsf{l})$ is identified by

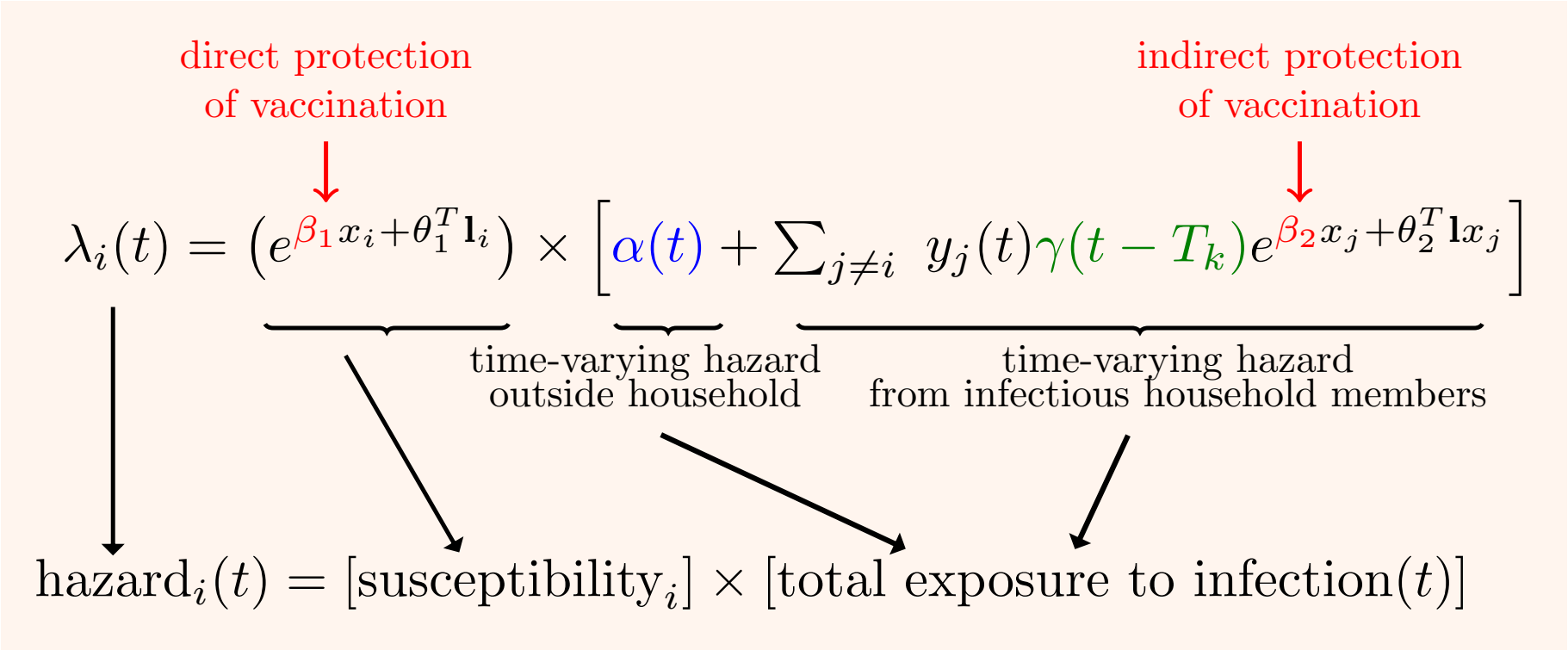
$$F_{I_i^j}(s|\mathsf{x},\mathsf{h}_{(i)},\mathsf{l}) = 1 - \exp\left[-\int_{t_{(i)}^j}^{t_{(i)}^j+s} \frac{f_i^j(u|\mathsf{x},\mathsf{h}_{(i)},\mathsf{l})}{S_i^j(u|\mathsf{x},\mathsf{h}_{(i)},\mathsf{l})}du\right] \text{ for } j=0,\ldots,n-1$$

Simulation: Estimations of causal estimands

Cluster	Treatment	Probability estimands				
		$\hat{CE}(t,0,0,1)$	$\hat{SE}(t,0)$	$\hat{IE}(t,0,0)$	$DE(t)$	$IDE(t)$
Constant external and internal hazards						
2	Obs.	0.005	-0.015	-0.036	-0.013	-0.036
	Bernoulli	0.004	-0.015	-0.036	-0.014	-0.038
	Block	0.004	-0.013	-0.036	0.025	-
	Cluster	0.004	-0.013	-0.035	-0.048	-
4	Obs.	0.026	-0.014	-0.084	-0.012	-0.073
	Bernoulli	0.025	-0.013	-0.082	-0.012	-0.063
	Block	0.026	-0.015	-0.082	0.016	-
	Cluster	0.025	-0.014	-0.083	-0.099	-
8	Obs.	0.068	-0.013	-0.131	-0.010	-0.088
	Bernoulli	0.069	-0.014	-0.133	-0.010	-0.096
	Block	0.069	-0.014	-0.132	0.010	-
	Cluster	0.070	-0.016	-0.132	-0.154	-

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.

Stochastic model for disease transmission



- $\alpha(t)$ is exogenous hazard of infection, $\gamma(t)$ is endogenous hazard of transmission between individuals
- β_1 is for susceptibility effect, β_2 is for infectiousness effect
- θ_1 and θ_2 are covariate effects of susceptibility and infectiousness

New vaccine estimands based on hazards

Controlled hazard ratio vaccine effects

- Susceptibility hazard ratio:

$$HSE^{C}(t, \mathrm{x}_{(i)}, \mathrm{h}_{(i)}, \mathrm{l}) = \frac{\lambda_i(t \mid 1, \mathrm{x}_{(i)}, \mathrm{h}_{(i)}, \mathrm{l})}{\lambda_i(t \mid 0, \mathrm{x}_{(i)}, \mathrm{h}_{(i)}, \mathrm{l})} = e^{\beta_1}$$

- Infectiousness hazard ratio:

$$HIE^{C}(t, h_j, h_j', \mathrm{x}_{(j)}, \mathrm{h}_{(i,j)}, \mathrm{l}) = \frac{\lambda_i(t|1,\mathrm{x}_{(j)},h_j',\mathrm{h}_{(i,j)},\mathrm{l}) - \lambda_i(t|1,\mathrm{x}_{(j)},h_j,\mathrm{h}_{(i,j)},\mathrm{l})}{\lambda_i(t|0,\mathrm{x}_{(j)},h_j',\mathrm{h}_{(i,j)},\mathrm{l}) - \lambda_i(t|0,\mathrm{x}_{(j)},h_j,\mathrm{h}_{(i,j)},\mathrm{l})} = e^{\beta_2}$$

  where $y_j(t) = 1$ as specified in h_j', and $y_j(t) = 0$ as specified in h_j.

- Contagion cumulative hazard ratio:

$$HCE^{C}(t; h_j'', h_j', \mathrm{h}_{(i,j)}, \mathrm{l}) = \frac{\int_0^t \left[\lambda_i(u;0,h_j'(t),\mathrm{h}_{(i,j)},\mathrm{l}) - \lambda_i(u;0,h_j(t),\mathrm{h}_{(i,j)},\mathrm{l})\right] du}{\int_0^t \left[\lambda_i(u;0,h_j''(t),\mathrm{h}_{(i,j)},\mathrm{l}) - \lambda_i(u;0,h_j(t),\mathrm{h}_{(i,j)},\mathrm{l})\right] du} = \frac{\int_{t_j'}^t \gamma(u) du}{\int_{t_j''}^t \gamma(u) du}$$

  where $y_j(t) = 1$ as specified in h_j', and $y_j(t) = 0$ as specified in h_j.

Simulation: Estimations of causal estimands

Cluster	Treatment	Hazard estimands		Probability estimands				
		$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{CE}$	$\hat{SE}$	$\hat{IE}$	$DE(t)$	$IDE(t)$
Constant external and internal hazards								
2	Obs.	-0.119	-2.271	0.005	-0.015	-0.036	-0.013	-0.036
	Bernoulli	-0.115	-2.334	0.004	-0.015	-0.036	-0.014	-0.038
	Block	-0.102	-2.364	0.004	-0.013	-0.036	0.025	-
	Cluster	-0.103	-2.288	0.004	-0.013	-0.035	-0.048	-
4	Obs.	-0.105	-2.368	0.026	-0.014	-0.084	-0.012	-0.073
	Bernoulli	-0.105	-2.286	0.025	-0.013	-0.082	-0.012	-0.063
	Block	-0.116	-2.278	0.026	-0.015	-0.082	0.016	-
	Cluster.	-0.107	-2.323	0.025	-0.014	-0.083	-0.099	-
8	Obs.	-0.100	-2.287	0.068	-0.013	-0.131	-0.010	-0.088
	Bernoulli	-0.106	-2.331	0.069	-0.014	-0.133	-0.010	-0.096
	Block	-0.111	-2.311	0.069	-0.014	-0.132	0.010	-
	Cluster	-0.120	-2.299	0.070	-0.016	-0.132	-0.154	-

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.

Summary

- We articulate the causal structure between individuals' treatments and outcomes in infectious disease, and illustrate the identification strategy for the potential outcomes under contagion, in the example of inter-connected clusters.
- A class of fundamental (controlled- and marginalized-) causal estimands for the susceptibility, infectiousness and contagion effect of vaccines are proposed, and comprehensively compared to popular estimands in contemporary epidemiology.
- We provide the identification of causal estimands non-parametrically, and further apply a generalized Cox-type transmission hazard model to facilitate the inference of causal estimands.
- We promote hazard ratio as alternative causal estimands for the susceptibility and infectiousness effect, and compared them to existing estimands for vaccine efficacy.

Acknowledgement

This work was supported by NIH grant 1DP2HD091799-01. Xiaoxuan Cai was supported by a fellowship from Takeda Pharmaceutical Company.
We thank Peter Aronow, Olga Morozova, Virginia Pitzer for their great suggestions.

xiaoxuan.cai@yale.edu
https://xiaoxuan-cai.github.io/