Causal inference for infectious disease intervention in inter-connected clusters

Xiaoxuan Cai

joint work with Eben Kenah, Forrest W. Crawford

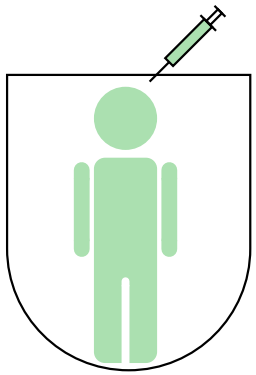
Department of Biostatistics, Yale School of Public Health

August 6, 2020 Joint Statistical Meeting

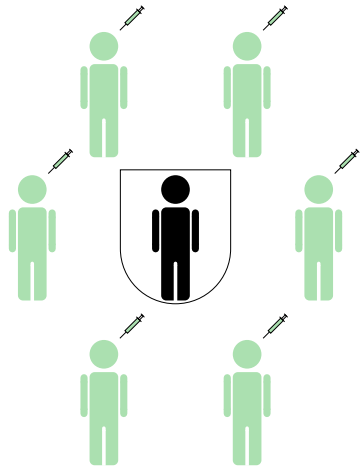
Infectious disease and vaccination

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, contagion effect, infectiousness effect...
- Vaccines for Polio, Influenza, HIV/AIDS, Malaria and etc.



Direct protection



Indirect protection

Why infectious disease is difficult to study?

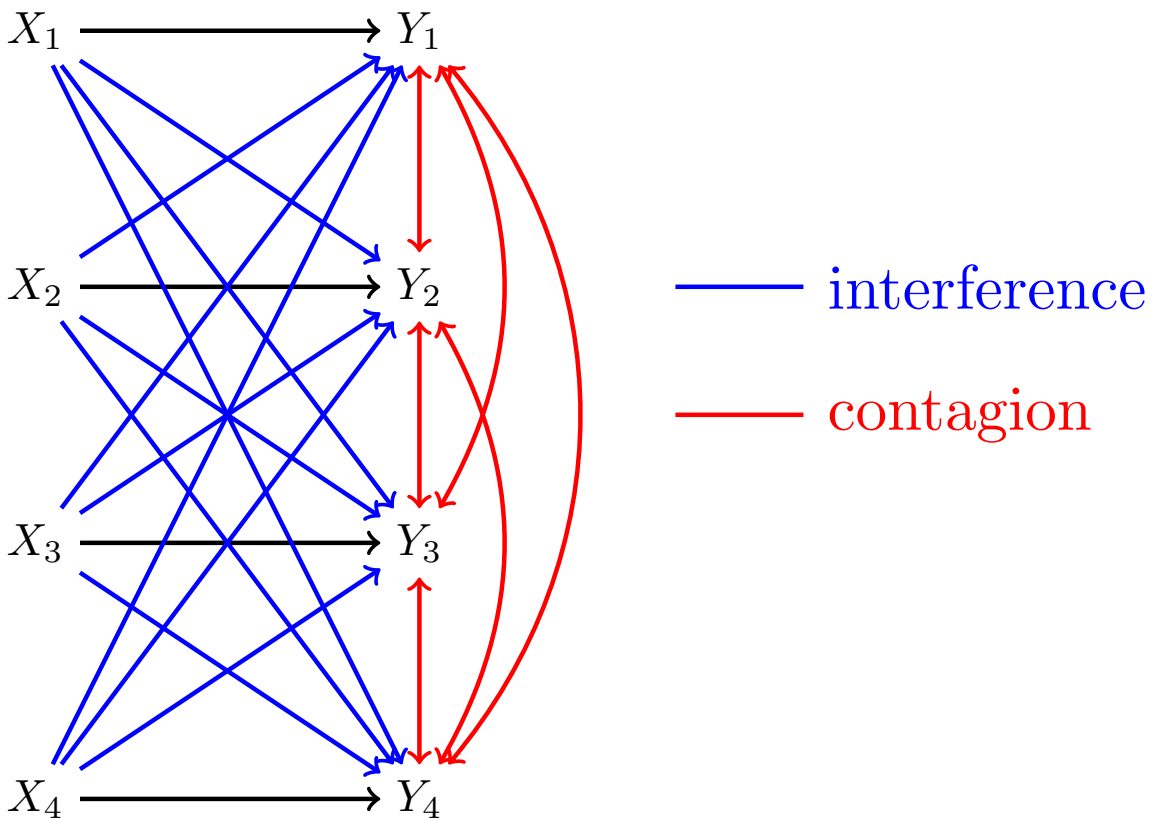
Research on transmission of infectious disease has some unique features and challenges.

- 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

Notation

Consider a cluster of n individuals, $i = 1, \ldots, n$

- Treatment assignment: $\mathrm{X} = (X_1, \ldots, X_n)$
- Infection time: $\mathrm{T} = (T_1, \ldots, T_n)$
- Infection outcome: $\mathrm{Y}(t) = (Y_1(t), \ldots, Y_n(t))$
  ,where $Y_i(t) = \mathbb{1}\{T_i < t\}$
- Others' treatments: $\mathrm{X}_{(i)} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$
- Others' infection times: $\mathrm{T}_{(i)} = (T_1, \ldots, T_{i-1}, T_{i+1}, \ldots, T_n)$
- Others' infection history: $\boldsymbol{\mathcal{H}}_{(i)} = \{Y_j(s) : s \geq 0, j \neq i\}$
  , or $\boldsymbol{\mathcal{H}}_{(i)} = \mathrm{T}_{(i)}$

Notation

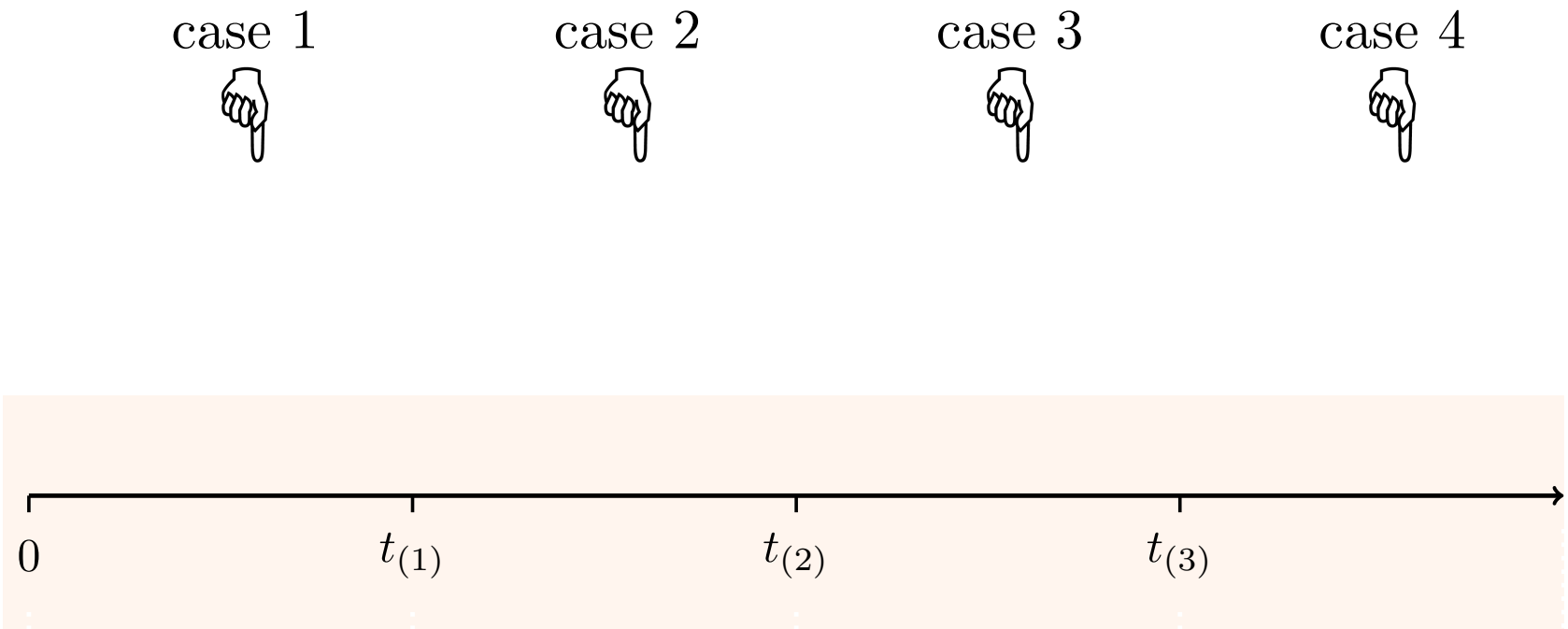
Define $T_i(\mathrm{X} = \mathrm{x}, \boldsymbol{\mathcal{H}}_{(i)} = \mathrm{h}_{(i)})$ and $Y_i(t; \mathrm{X} = \mathrm{x}, \boldsymbol{\mathcal{H}}_{(i)} = \mathrm{h}_{(i)})$ as the counterfactual infection time and outcome of i under a joint treatment x and a deterministic infection history $\mathrm{h}_{(i)}$ of other individuals, respectively.

Goal 1

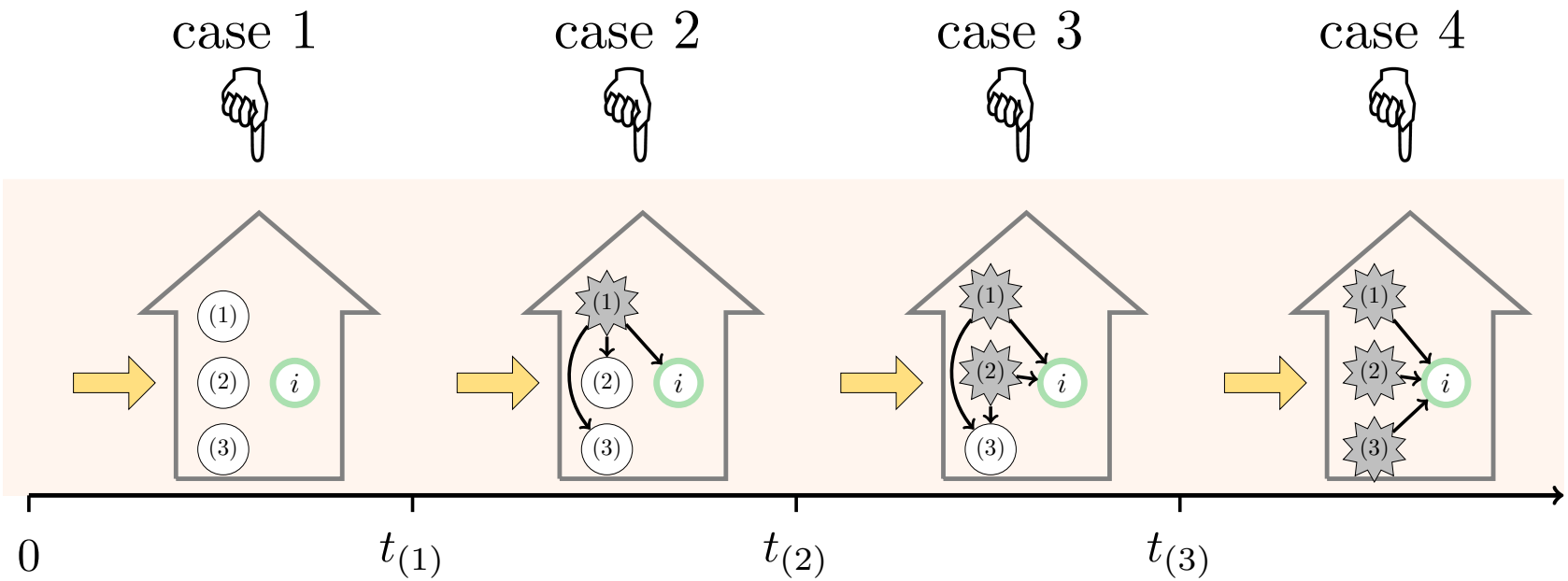
Identify $T_i(\mathrm{h}_{(i)}, \mathrm{x})$ or $Y_i(t; \mathrm{h}_{(i)}, \mathrm{x})$ under joint intervention $(\mathrm{h}_{(i)}, \mathrm{x})$.

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

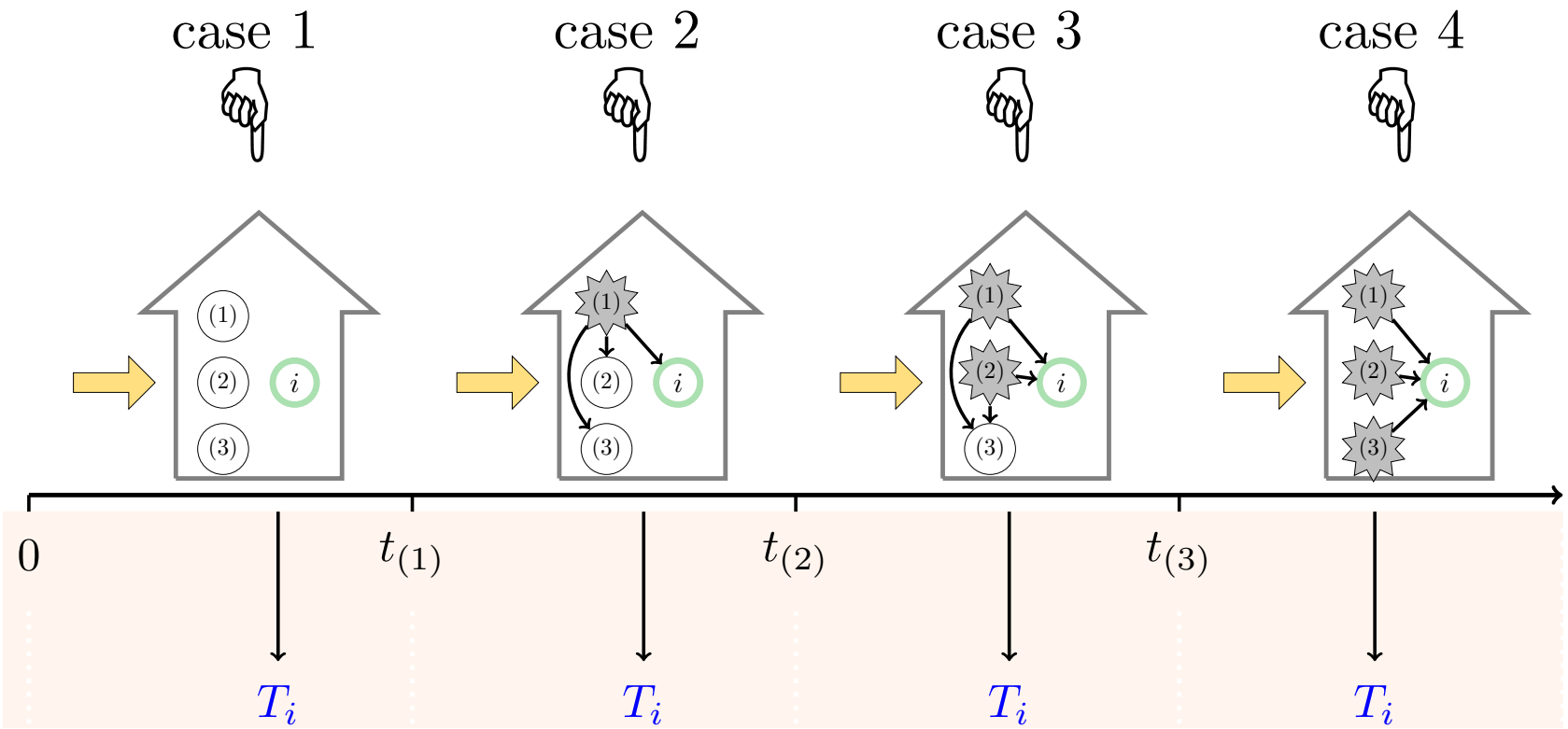
Clusters of 4 individuals



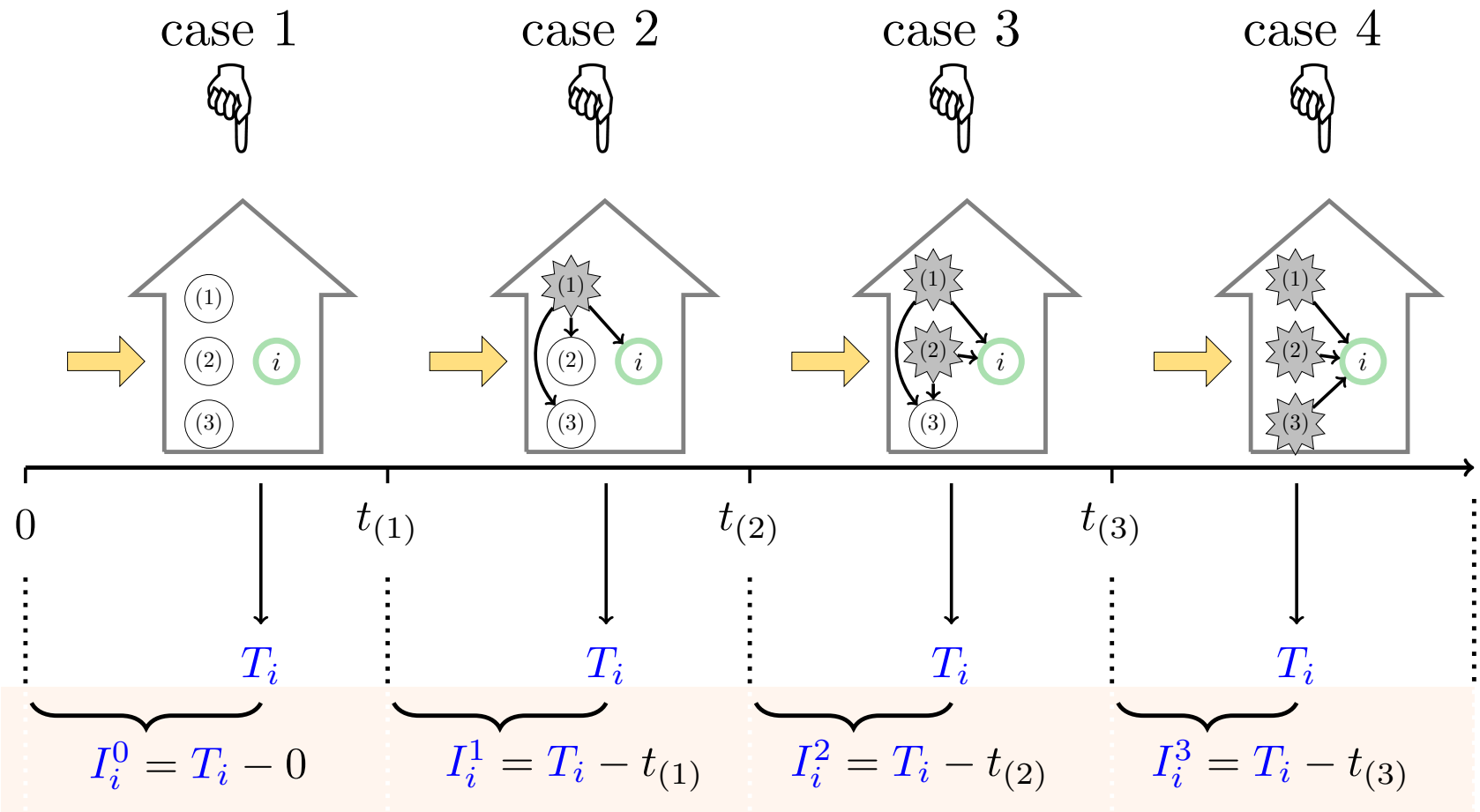
Clusters of 4 individuals



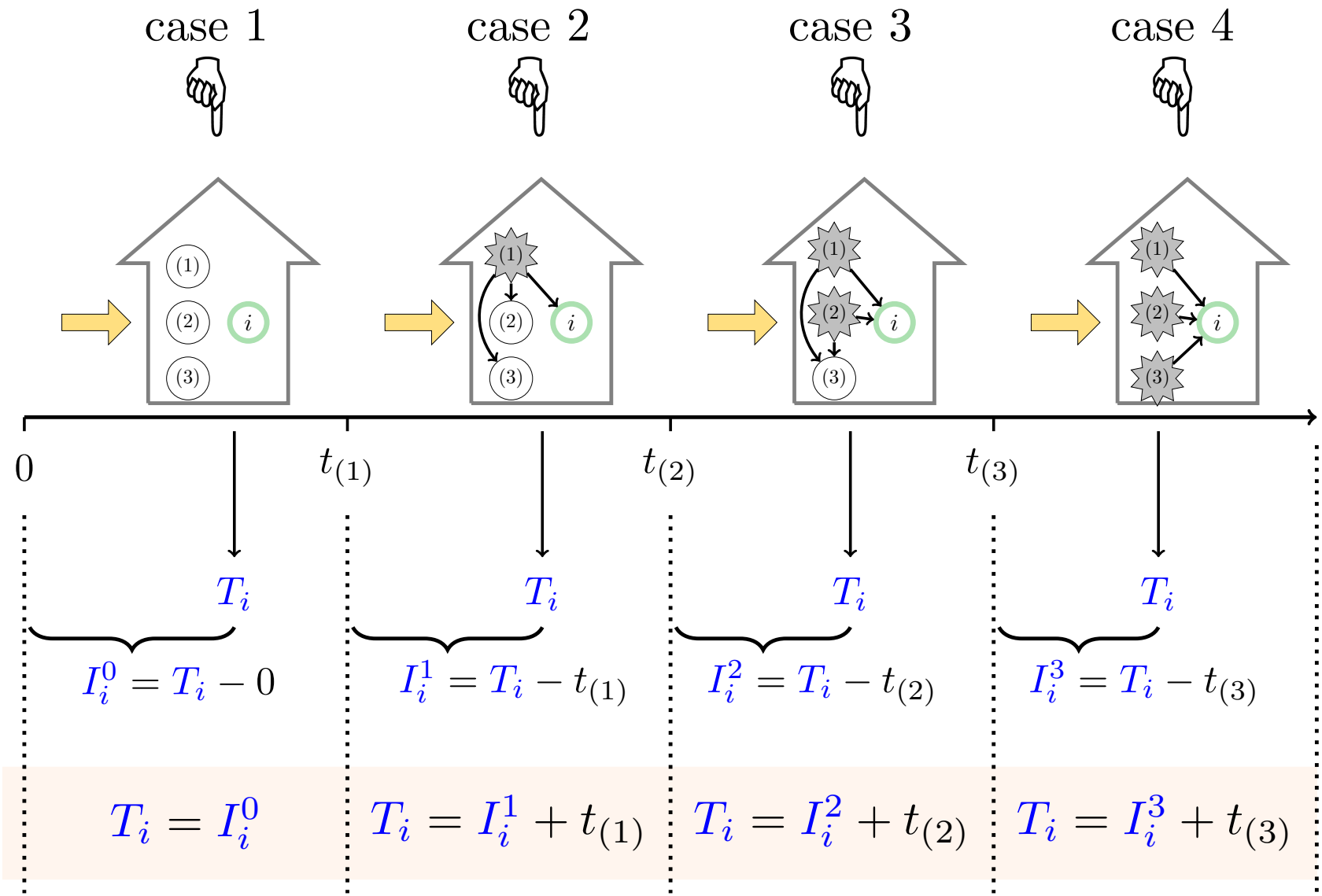
Clusters of 4 individuals



Clusters of 4 individuals



Clusters of 4 individuals



Decomposition of $T_i(\mathrm{x}, \mathrm{h}_{(i)})$

For a deterministic $\mathrm{h}_{(i)}$ with $\mathrm{t}_{(i)}$, reorder infection times as $\mathrm{t}_{(i)} = (t^1_{(i)}, t^2_{(i)}, \ldots, t^{n-1}_{(i)})$, then we rewrite

$$
\begin{aligned}
&T_i(\mathrm{x}, \mathrm{h}_{(i)}) \\
&= \begin{cases}
I_i^0(\mathrm{x}, \mathrm{h}_{(i)}) & \text{if } I_i^0(\mathrm{x}, \mathrm{h}_{(i)}) < t^1_{(i)} \\
t^1_{(i)} + I_i^1(\mathrm{x}, \mathrm{h}_{(i)}) & \text{if } I_i^0(\mathrm{x}, \mathrm{h}_{(i)}) \geq t^1_{(i)}, I_i^1(\mathrm{x}, \mathrm{h}_{(i)}) < t^2_{(i)} - t^1_{(i)} \\
\vdots & \vdots \\
t^{n-1}_{(i)} + I_i^{n-1}(\mathrm{x}, \mathrm{h}_{(i)}) & \text{if } I_i^0(\mathrm{x}, \mathrm{h}_{(i)}) \geq t^1_{(i)}, \ldots, I_i^{n-2}(\mathrm{x}, \mathrm{h}_{(i)}) \geq t^{n-1}_{(i)} - t^{n-2}_{(i)}
\end{cases}
\end{aligned}
$$

where $I_i^k(\mathrm{x}, \mathrm{h}_{(i)})$ is the potential additional time to infection after the k^{th} infection at $t^k_{(i)}$.

Main Result: Exposure-controlled potential outcome

Theorem: Identification of exposure-controlled potential outcomes

Under conventional assumptions in causal inference,

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

where $F_{I_i^j}(s \,|\, \mathsf{x}, \mathsf{h}_{(i)}, \mathsf{l})$ is the distribution of the latent waiting times to infection $I_i^k(\mathsf{x}, \mathsf{h}_{(i)})$ and is identified by

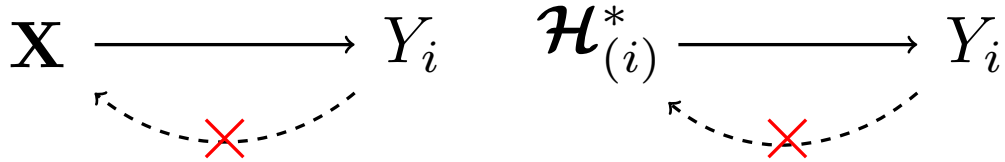
$$F_{I_i^j}(s|\mathsf{x}, \mathsf{h}_{(i)}, \mathsf{l}) = 1 - \exp\left[- \int_{t_{(i)}^j}^{t_{(i)}^j + s} \frac{f_i^j(u|\mathsf{x}, \mathsf{h}_{(i)}, \mathsf{l})}{S_i^j(u|\mathsf{x}, \mathsf{h}_{(i)}, \mathsf{l})} du \right] \text{ for } j = 0, \ldots, n-1$$

where $f_i^j(u|\mathsf{x}, \mathsf{h}_{(i)}, \mathsf{l})$ and $S_i^j(u|\mathsf{x}, \mathsf{h}_{(i)}, \mathsf{l})$ are identifiable by standard results from competing risks.

The natural distribution of the deterministic $h_{(i)}$

Consider a cluster of n individuals, $i = 1, \ldots, n$.

- $\mathcal{H}^*_{(i)}(x)$: the random history of infection times in individuals other than i under $X = x$, in an otherwise identical group of $n - 1$ individuals in which i is absent, or cannot transmit infection.
- $\mathcal{H}^*_{(i)}(x)$ **is equivalent to** $\mathcal{H}_{(i)}(x_{(i)})$ **in distribution for** i**'s infection.**



Define $Y_i(t; x, \mathcal{H}^*_{(i)}(x_{(i)}))$ as the counterfactual infection outcome i under a joint treatment x and the natural distribution of $h_{(i)}$ under $x_{(i)}$.

Goal 2

Identify $Y_i\left(t; x, \mathcal{H}^*_{(i)}(x_{(i)})\right)$ under joint intervention x and $\mathcal{H}^*_{(i)}(x_{(i)})$.

Main Result: Exposure-marginalized (Natural) potential outcomes

Theorem: Identification of exposure-marginalized potential outcomes

Under conventional assumptions in causal inference,

$$\mathbb{E}\left[Y_i\left(t; x_i, \mathsf{x}_{(i)}, \boldsymbol{\mathcal{H}}^*_{(i)}(\mathsf{x}_{(i)})\right)|\mathsf{L}=\mathsf{l}\right] = \int \mathbb{E}[Y_i(t; x_i, \mathsf{x}_{(i)}, \mathsf{h}_{(i)})|\mathsf{L}=\mathsf{l}]\,\mathrm{d}G^*_{(i)}(\mathsf{h}_{(i)}|\mathsf{x}_{(i)}, \mathsf{l}_{(i)})$$

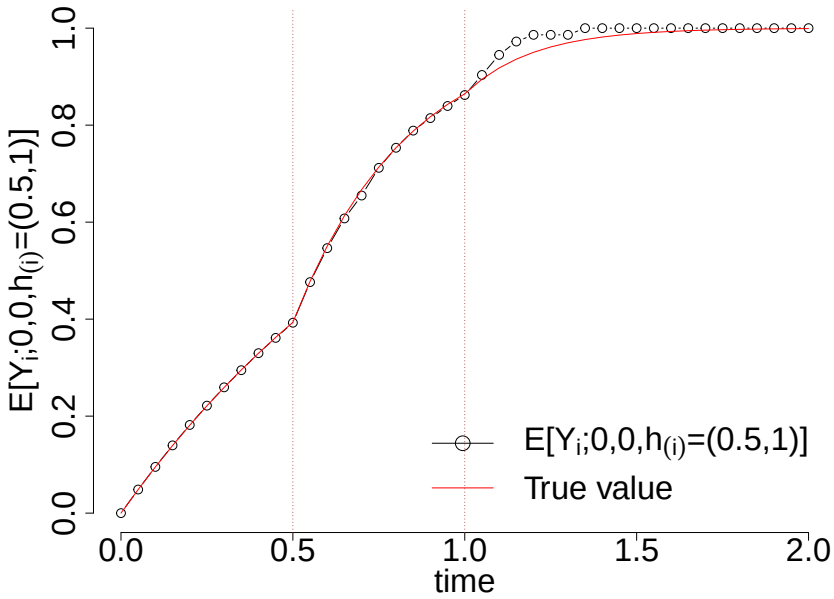
where $G^*_{(i)}(\mathsf{h}_{(i)}|\mathsf{x}_{(i)}, \mathsf{l}_{(i)})$ is the distribution of the potential infection history $\boldsymbol{\mathcal{H}}^*_{(i)}(\mathsf{x}_{(i)})$ given $\mathsf{L}_{(i)} = \mathsf{l}_{(i)}$ and can be identified by

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

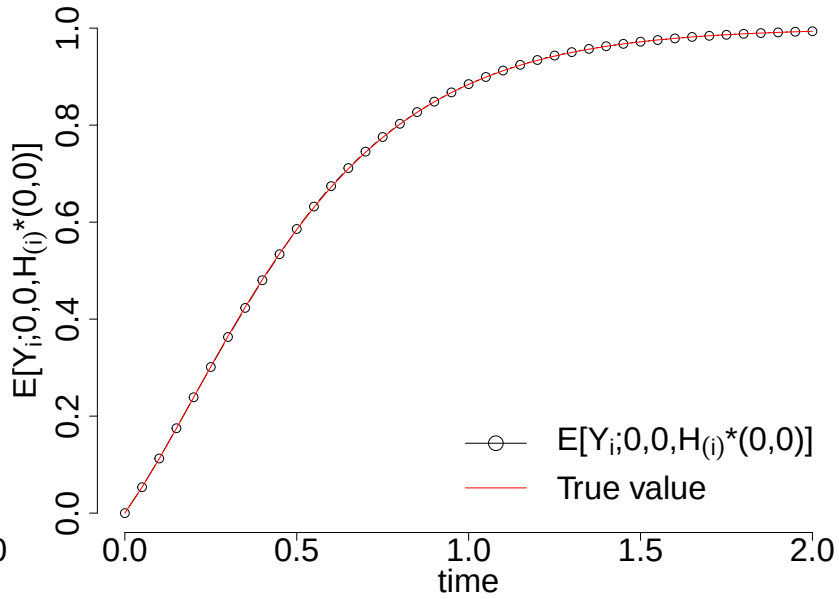
where $f^j_i(u|\mathsf{x}, \mathsf{h}_{(i)}, \mathsf{l})$ and $S^j_i(u|\mathsf{x}, \mathsf{h}_{(i)}, \mathsf{l})$ are identifiable by standard results from competing risks.

Simulation: causal identification for potential outcomes

We simulate N=100,000 clusters of **three** individuals with constant exogenous and internal infection hazards, without covariates.



Controlled-exposure outcome



Marginalized-exposure outcome

Causal estimands: Exposure-controlled causal estimands

Exposure-controlled causal estimands

- Susceptibility effect

$$SE_i(t, \mathrm{x}_{(i)}, \mathrm{h}_{(i)}) = \mathbb{E}[Y_i(t; x_i = 1, \mathrm{x}_{(i)}, \mathrm{h}_{(i)}) - Y_i(t; x_i = 0, \mathrm{x}_{(i)}, \mathrm{h}_{(i)})]$$

- Infectiousness effect

$$IE_i(t, x_i, \mathrm{h}_{(i)}) = \mathbb{E}[Y_i(t; x_i, \mathrm{x}_{(i)} = 1, \mathrm{h}_{(i)}) - Y_i(t; x_i, \mathrm{x}_{(i)} = 0, \mathrm{h}_{(i)})]$$

- Contagion effect

$$CE_i(t, \mathrm{x}, \mathrm{h}_{(i)}, \mathrm{h}'_{(i)}) = \mathbb{E}[Y_i(t; \mathrm{x}, \mathrm{h}_{(i)}) - Y_i(t; \mathrm{x}, \mathrm{h}'_{(i)})]$$

- 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

Causal estimands: Exposure-marginalized causal estimands

Exposure-marginalized (natural) causal estimands

- Susceptibility effect

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

- Infectiousness effect

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

- Contagion effect

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

Other commonly used estimators:

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

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

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.

Cox-type hazard model for pairwise infection times

Assume previously infected individuals, along with a exogenous source of infection, impose **indepdendent** and **competing** risks of disease transmission to the remaining uninfected individuals.
For all i and all infected individual j, $j \neq i$, we consider a Cox-type hazard model for the pairwise infection times:

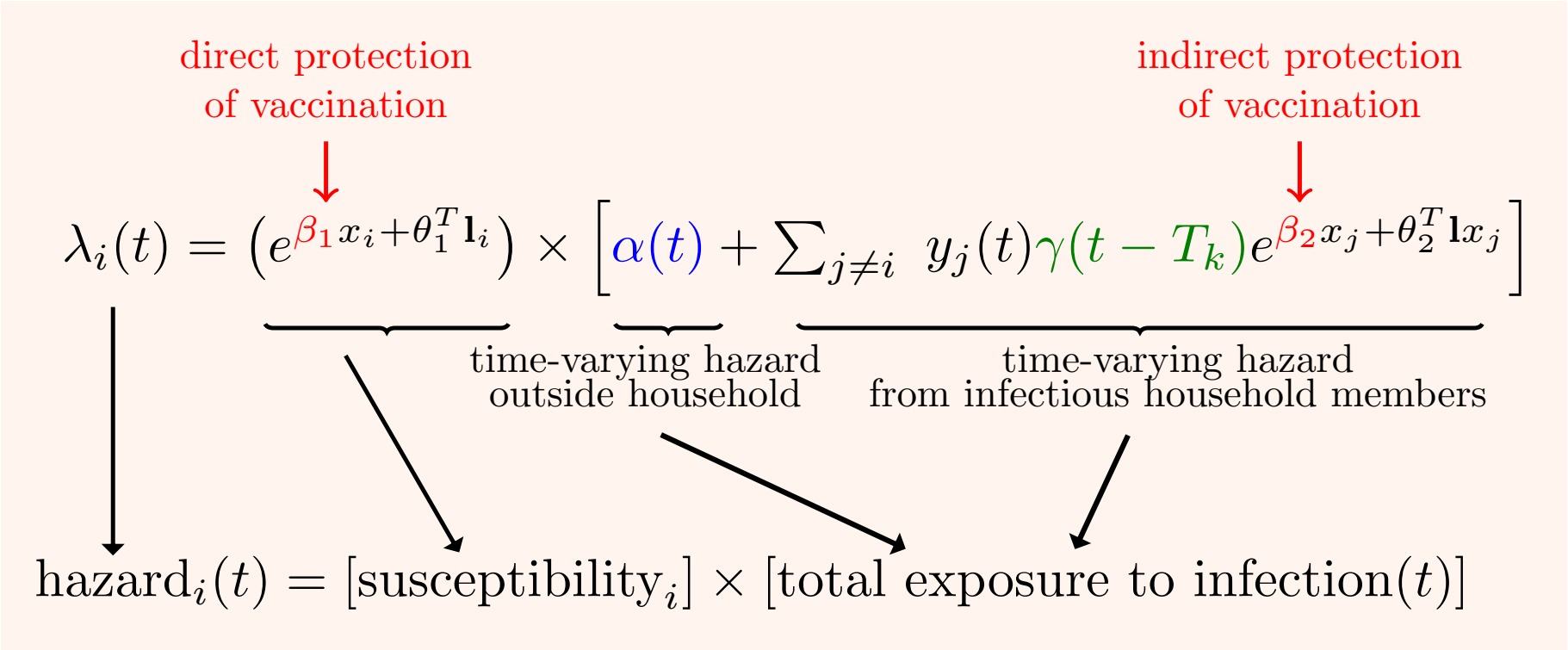
- **External** source of infection:

$$\lambda_{0i}(t \mid x_i, \mathsf{I}_i) = \alpha(t) \exp[\beta_1 x_i + \theta_1^T \mathsf{I}_i]$$

- **Internal** source of infection from infectious j to yet-uninfected i:

$$\lambda_{ji}(t \mid x_i, \mathsf{I}_i) = \gamma(t - t_j) \exp[\beta_1 x_i + \beta_2 x_j + \theta_1^T \mathsf{I}_i + \theta_2^T \mathsf{I}_j]$$

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/