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

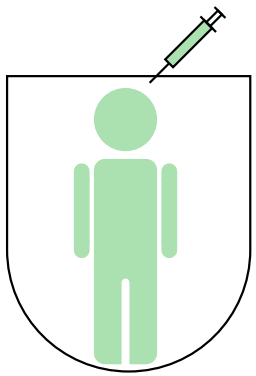
November 4, 2020

https://arxiv.org/abs/1912.04151

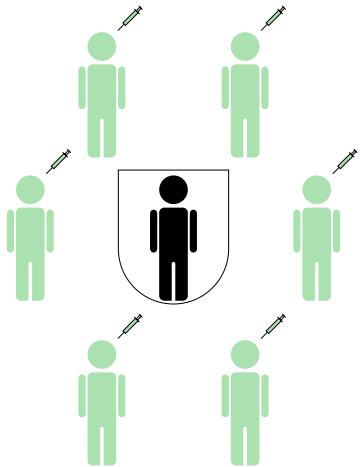
Infectious disease and vaccination

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
  Earlier exposure to infectious individual (higher "exposure to infection") increases the risk of infection.

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

Bias due to differential "exposure to infection"

Direct comparisons of treated and untreated individuals may not be valid due to differential "exposure to infection" [1,2,3].

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

- For example, if vaccinated individuals get infected later in general, then later infected and vaccinated subjects face higher exposure to infection, comparing to earlier infected and unvaccinated individuals. $\rightarrow$ not a fair comparison !

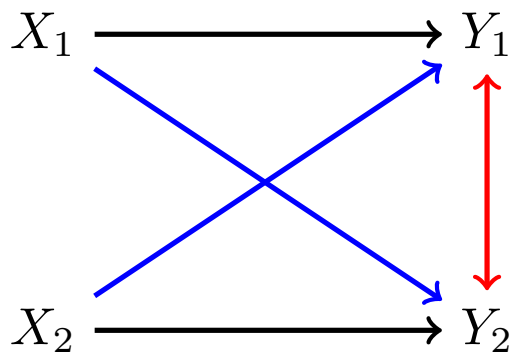
Can randomization solve the problem?

- No. Differential "exposure to infection" due to others' infection happens after randomization.

[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.

Symmetric partnership models

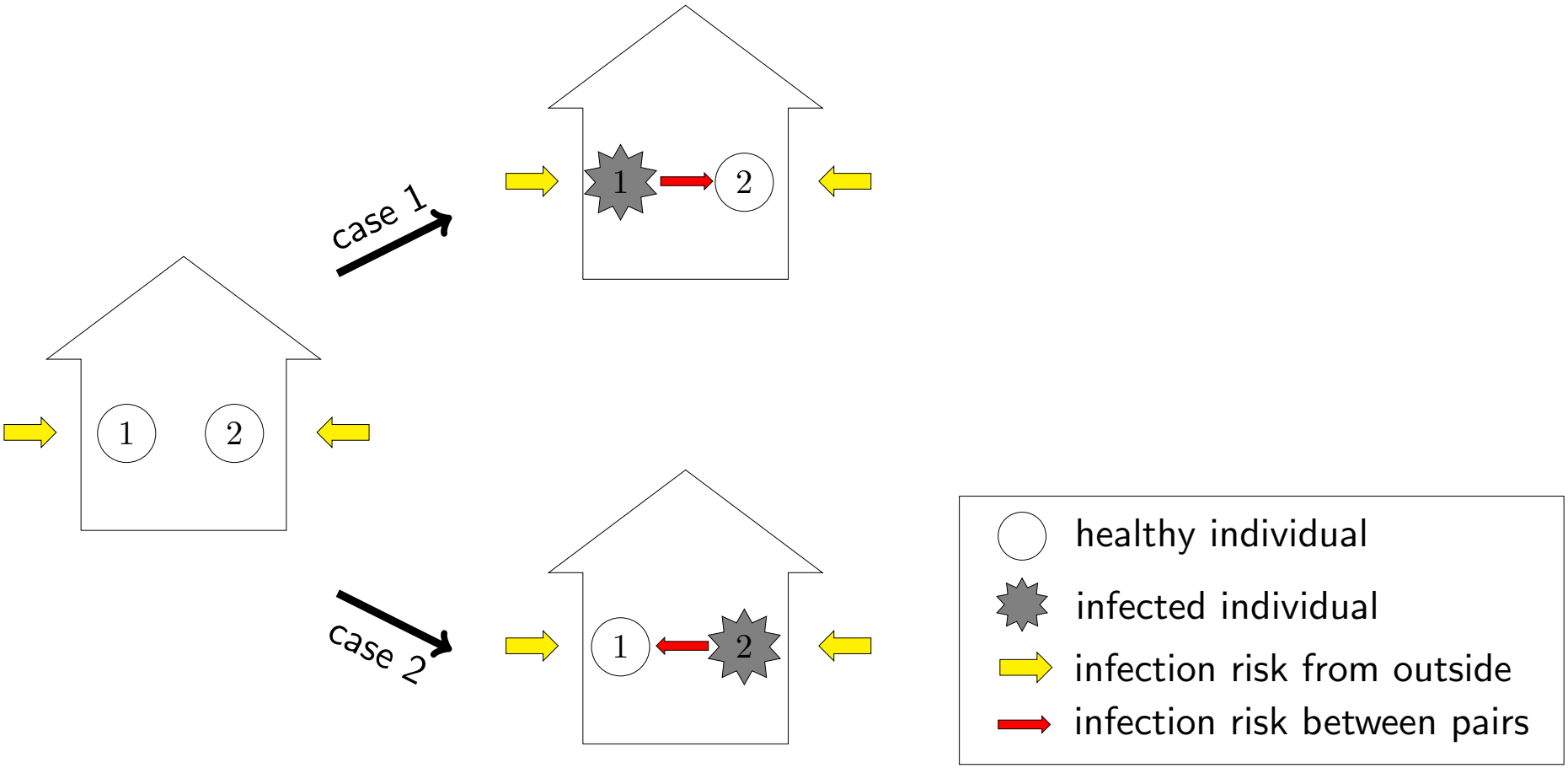
Partnership models have been widely understood as a useful framework to clarify causal relationship in epidemiology, and lay the foundation for more complex settings.
Consider two individuals with treatment X_1 and X_2 and infection outcome Y_1 and Y_2.



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

Two possibilities of the process

Separate the process into two distinctive possibilities: case 1 and case 2.



Unique challenges for causal identification under contagion

Problem: Differentiating exposure to infection, even 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:

Remove the bidirectional arrow by breaking down the transmission process into exclusive possibilities.

Notation

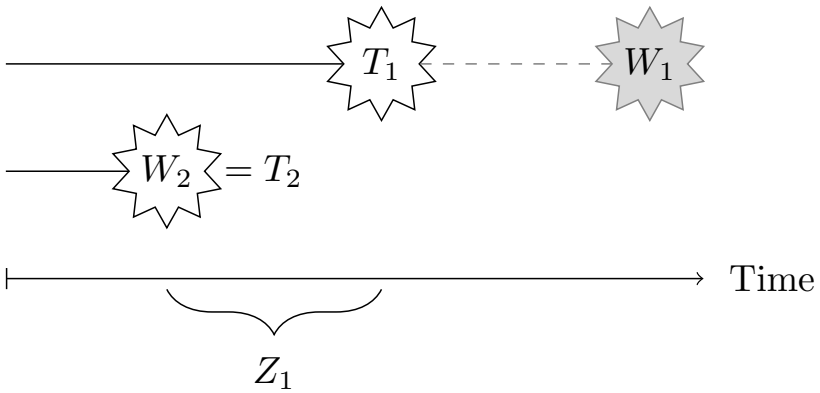
For the symmetric partnership models, consider individual 1 and 2 and let,

- Treatment assignment: $\mathbf{X} = (X_1, X_2)$
- Infection time: T_i for $i = 1, 2$
- Infection outcome: $Y_i(t) \equiv \mathbb{1}\{t \geq T_i\}$ for $i = 1, 2$

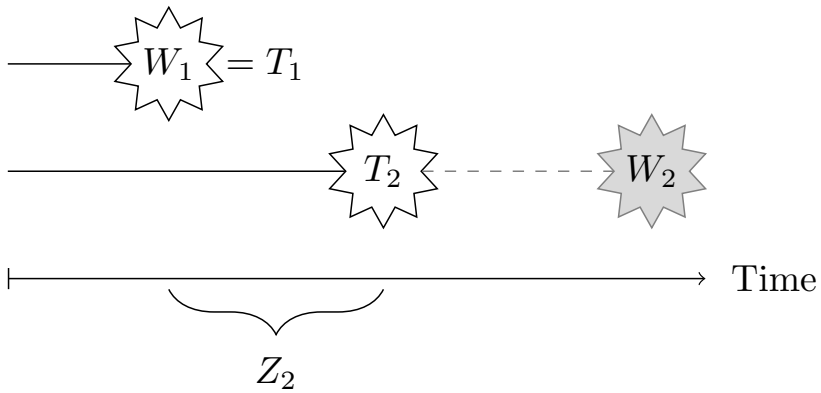
Instead of using the same T_i for both cases, introduce additional variables to distinguish two cases.

- Isolated infection time: W_i for $i = 1, 2$
- Extra infection time after partner's infection: $Z_i = T_i - W_j$ for $i \neq j$

Structure: Relationships between W_i, Z_i and T_i



Case 1: subject 2 gets infected first
T_1, Z_1 and $W_2 = T_2$ observed
W_1 censored, Z_2 undefined

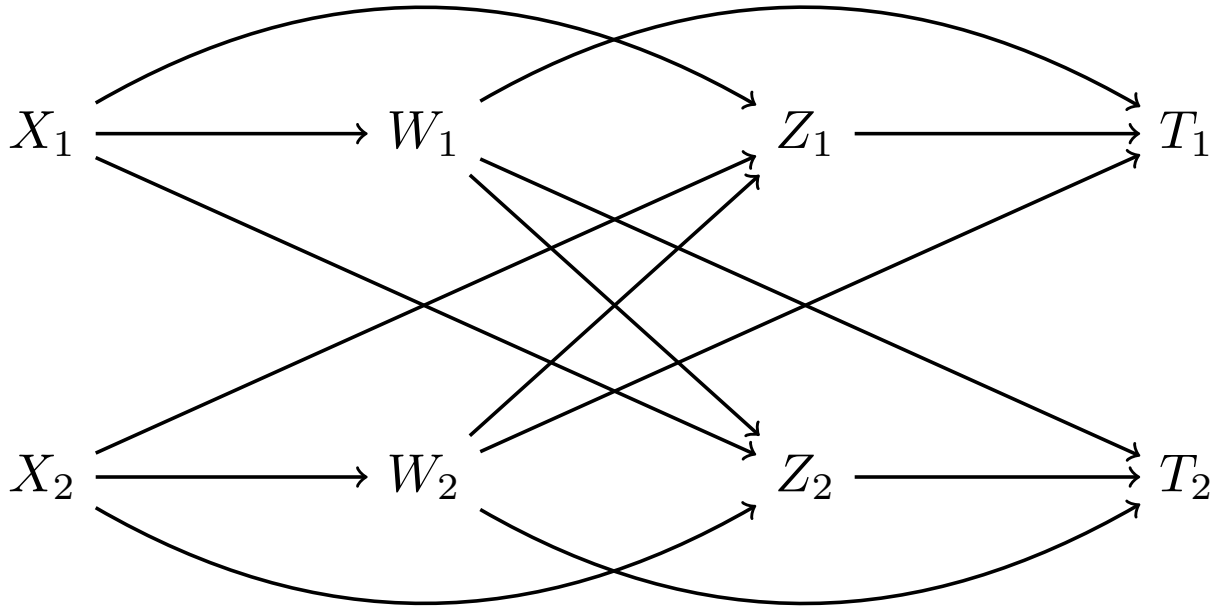


Case 2: subject 1 gets infected first
T_2, Z_2 and $W_1 = T_1$ observed
W_2 censored, Z_1 undefined

- W_i: spontaneous infection time only by external risk
- Z_i: additional time to infection after their partner's infection
- T_i: infection time of i

$$T_i = \begin{cases} W_i & \text{if } W_i < W_j \\ W_j + Z_i & \text{otherwise} \end{cases}$$

Graphical representation

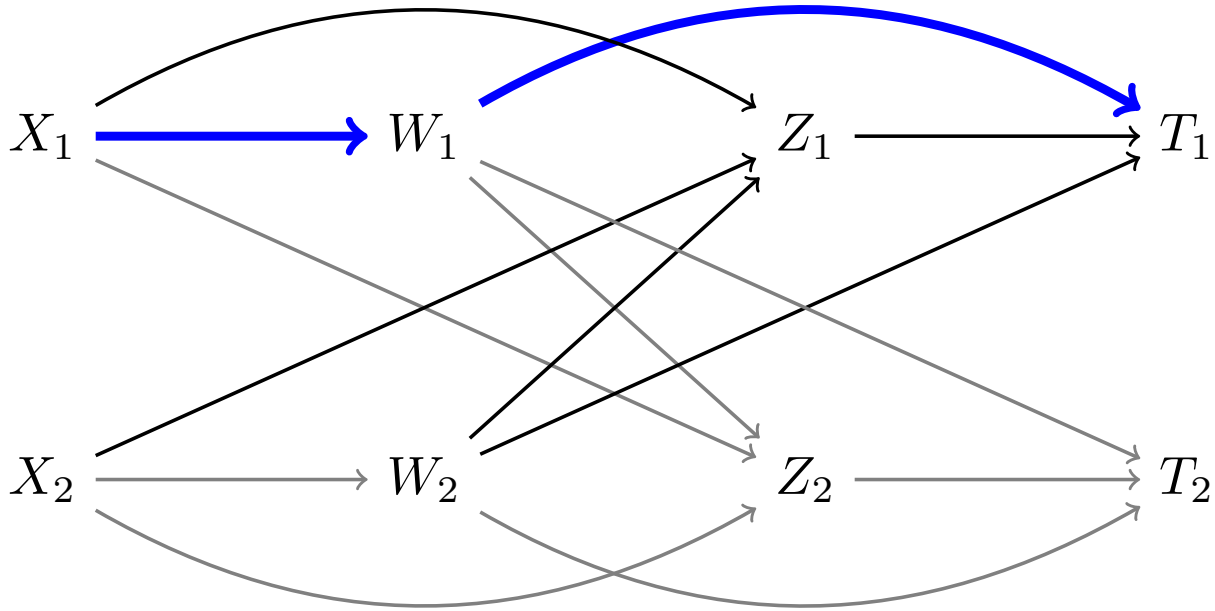


(Covariates L are omitted for simplicity)

- W_i: spontaneous infection time only by external risk
- Z_i: additional time to infection after their partner's infection
- T_i: infection time of i

Graphical representation

Use individual 1 as an example,

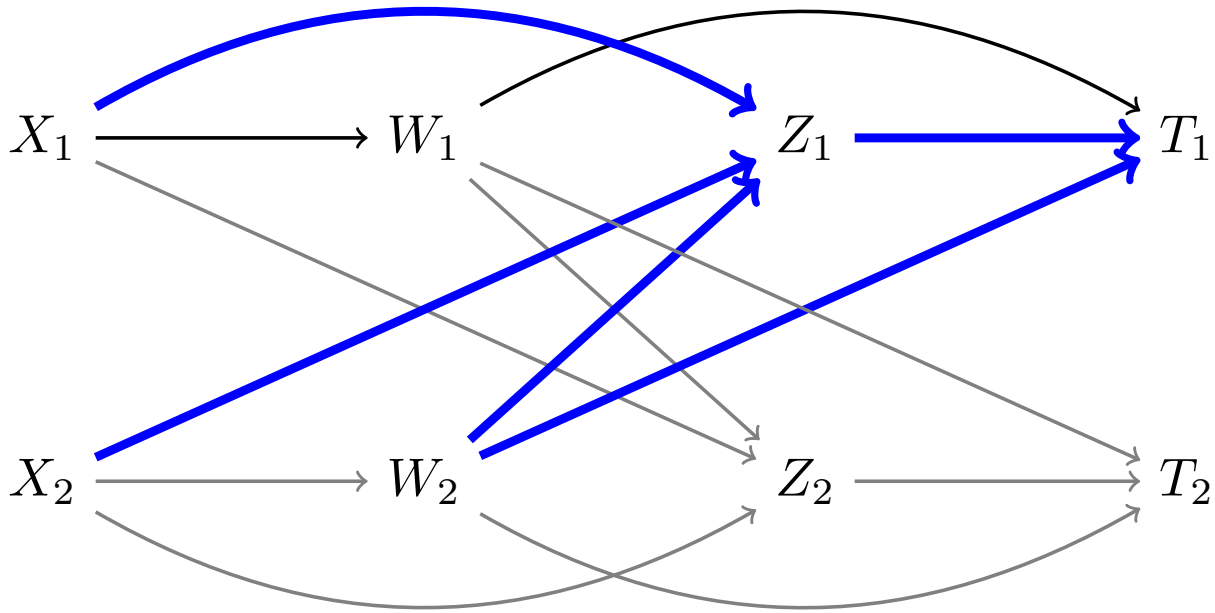


(Covariates L are omitted for simplicity)

- W_i: spontaneous infection time only by external risk
- Z_i: additional time to infection after their partner's infection
- T_i: infection time of i

Graphical representation

Use individual 1 as an example,



(Covariates L are omitted for simplicity)

- W_i: spontaneous infection time only by external risk
- Z_i: additional time to infection after their partner's infection
- T_i: infection time of i

Notation

Define counterfactual infection outcome $Y_i(t; s, x_i, x_j)$ for $i = 1, 2$ where $j \neq i$, when we fix (i) its **own treatment** $X_i = x_i$, (ii) **partner's treatment** $X_j = x_j$, and (iii) **partner's infection time** $W_j = s$.
Note: partner's infection time is treated as another type of intervention for the potential outcomes – "exposure to infection".

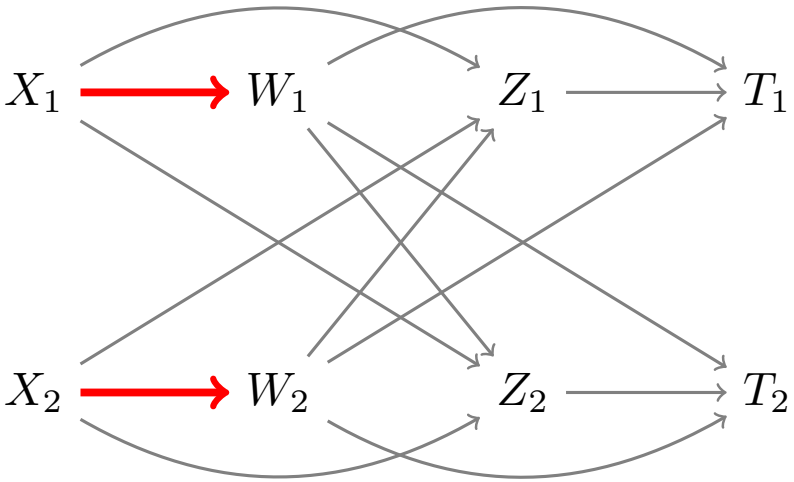
Goal:

Identify $Y_i(t; s, x_i, x_j)$, or equivalently $T_i(s, x_i, x_j)$, under the joint intervention (s, x_i, x_j) for $i = 1, 2$.

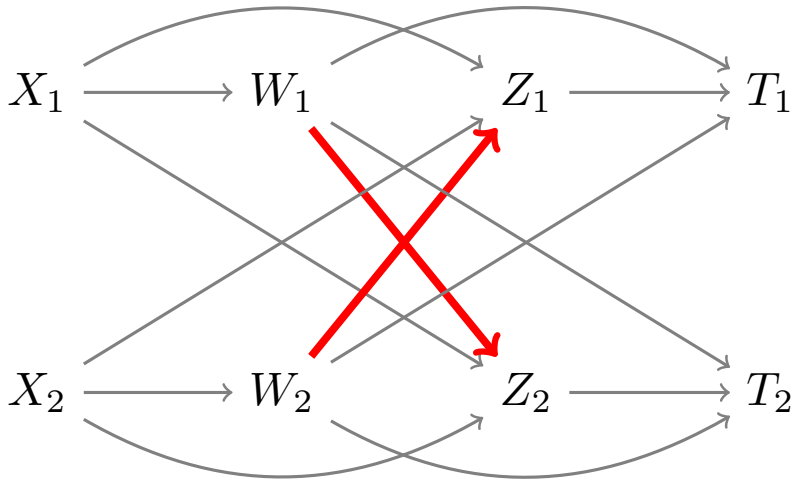
$$T_i(s, x_i, x_j) = \begin{cases} W_i(x_i) & \text{if } W_i(x_i) < W_j(x_j) \\ W_j(x_j) + Z_i(w_j(x_j); x_i, x_j) & \text{otherwise} \end{cases}$$

Main Result: Assumptions

- Treatment exchangeability: $Y_i(t; s, \mathrm{x}) \perp\!\!\!\perp \mathrm{X} \mid L$
- Consistency: $Y_i(t) = Y_i(t; s, \mathrm{x})$ when $\mathrm{X} = \mathrm{x}$ and $W_j = s$, $j \neq i$
- Positivity: $P(W_i = s, \mathrm{X} = \mathrm{x} | L = l) \in (0, 1)$ for all $x \in \mathcal{X}$, $l \in \mathcal{L}$ and $w_i > 0$ for $i = 1, 2$
- Infection independence: $W_i(x_i) \perp\!\!\!\perp W_j(x_j) | \mathrm{X}, L$ for $i \neq j$
- Infection ignorability: $Z_i(t; s, \mathrm{x}) \perp\!\!\!\perp W_j(x_j) | \mathrm{X}, L$ for $i \neq j$



Infecntion independence



Infecntion ignorability

Main Result: Identification

Identification theorem

Suppose Assumptions 1-5 hold. For fixed value of s, t and $\mathrm{x} = (x_i, x_j)$, 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:

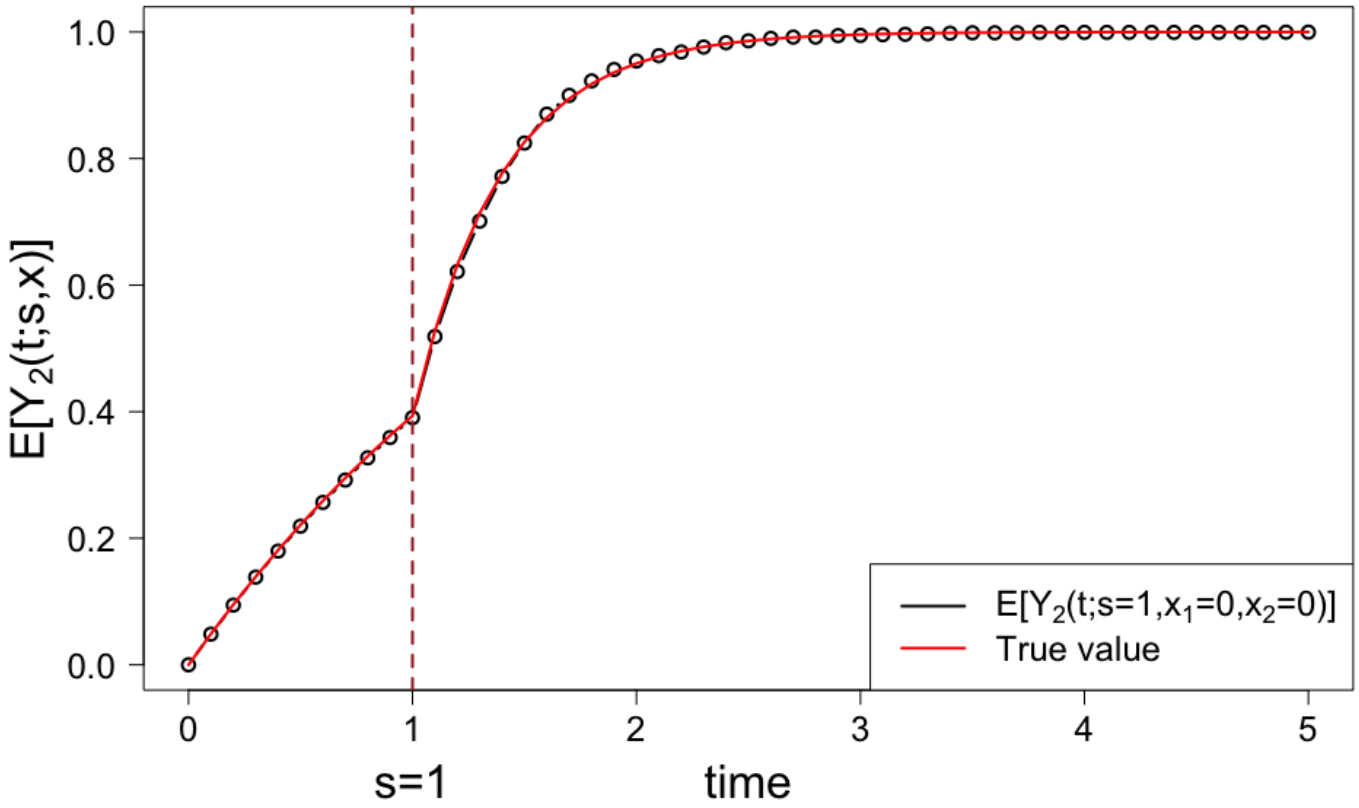
$$\mathbb{E}[Y_i(t; s, \mathrm{x})] = 1 \cdot p_i(s|\mathrm{x}) + \mathbb{E}[Y_i(t)|\, T_i \geq s, T_j = s, \mathrm{X} = \mathrm{x}] \cdot [1 - p_i(s|\mathrm{x})]$$

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

Note: Instead of using binary infection outcome by the end of observation, this causal identification is built on observation of infection time, which provides sufficient control for exposure to infection.

Simulation: causal identification for potential outcome

We simulate N=100,000 partnerships with contant exogenous hazard 0.5 and within-pair hazard 2. Vaccinations both decrease risks for vaccinated and reduce transmissibility by 50%.



Causal estimands: Exposure-controlled causal estimands

Exposure-controlled causal estimands

- Susceptibility effect ($s > 0$)

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

- Infectiousness effect ($s > 0$)

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

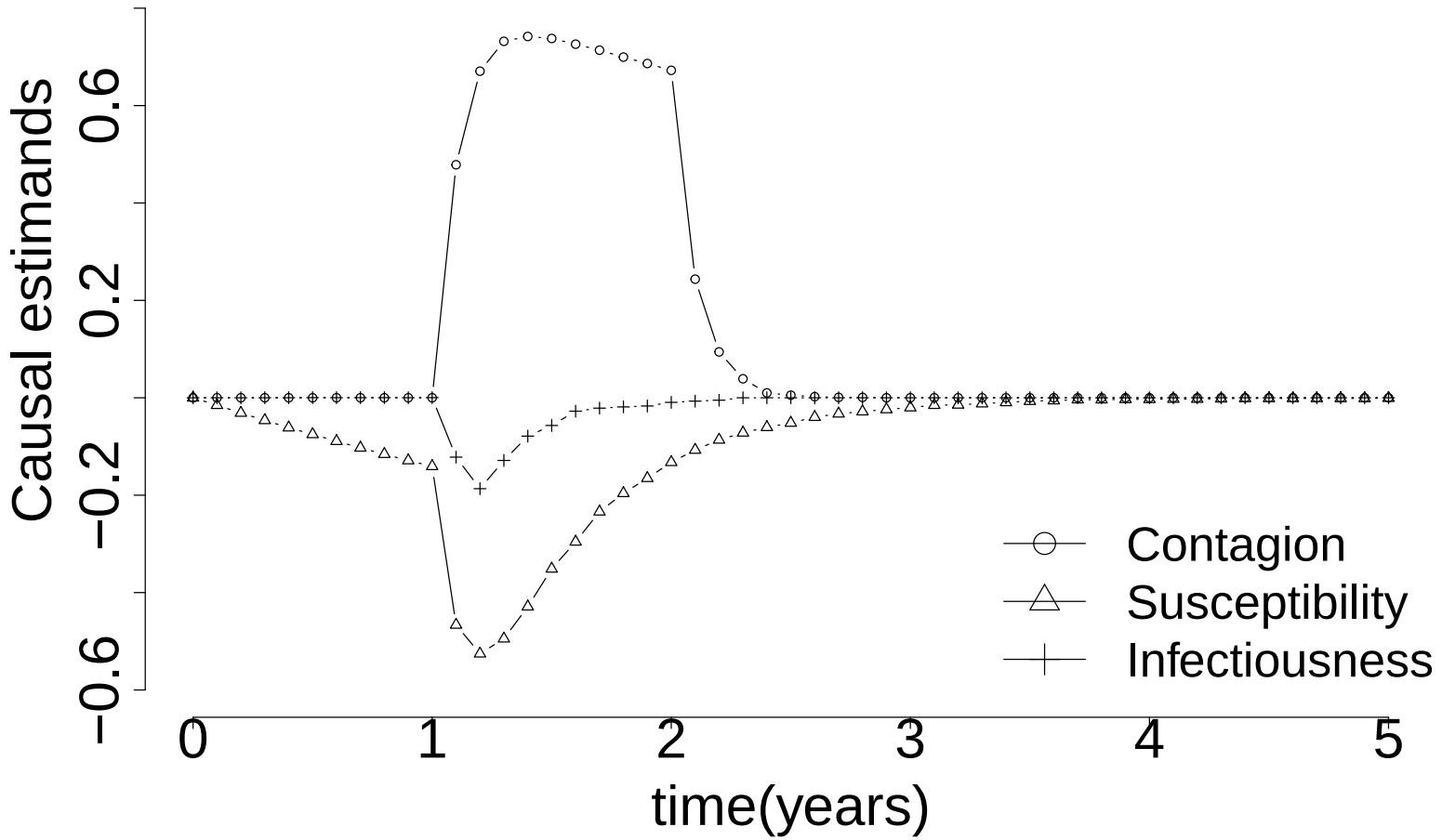
- Contagion effect ($s \neq s'$ and $\mathbf{X} = (0, 0)$)

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

- 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 constant exogenous hazard 0.5 and within-pair hazard 2. Vaccinations decrease infection risk of treated individuals and transmission ability both by 50%. We choose $s = 1$ and $s' = 2$.



Causal estimands: Exposure-marginalized causal estimands

Identification for natural potential outcomes

$$\mathbb{E}[Y_i(t; W_j(x_j'), \mathrm{x})|\mathrm{L} = \mathrm{l}] = \int_0^t \mathbb{E}[Y_i(t; w_j, \mathrm{x})|\mathrm{L} = \mathrm{l}] dF_j(w_j|x_j', \mathrm{l}_j).$$

Exposure-marginalized (natural) causal estimands

- Susceptibility effect

$$SE_i(t, x_j) = \mathbb{E}\big[Y_i\big(t; W_j(x_j), x_i = 1, x_j\big) - Y_i\big(t; W_j(x_j), x_i = 0, x_j\big)\big]$$

- Infectiousness effect

$$IE_i(t, x_i) = \mathbb{E}\big[Y_i\big(t; W_j(0), x_i, x_j = 1\big) - Y_i\big(t; W_j(0), x_i, x_j = 0\big)\big]$$

- Contagion effect

$$CE_i(t, \mathrm{x}) = \mathbb{E}\big[Y_i\big(t; W_j(0), x_i, x_j\big) - Y_i\big(t; W_j(1), x_i, x_j\big)\big]$$

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

Simulation: Estimations of natural causal estimands

Disease transmission dynamic:

- Constant external and internal hazards
- time-varying external and internal hazards

Treatment assignment:

- Bernoulli, Block, and Cluster randomization
- Observational studies with correlated (X_i, X_j) due to confounders

Natural causal estimands to be compared:

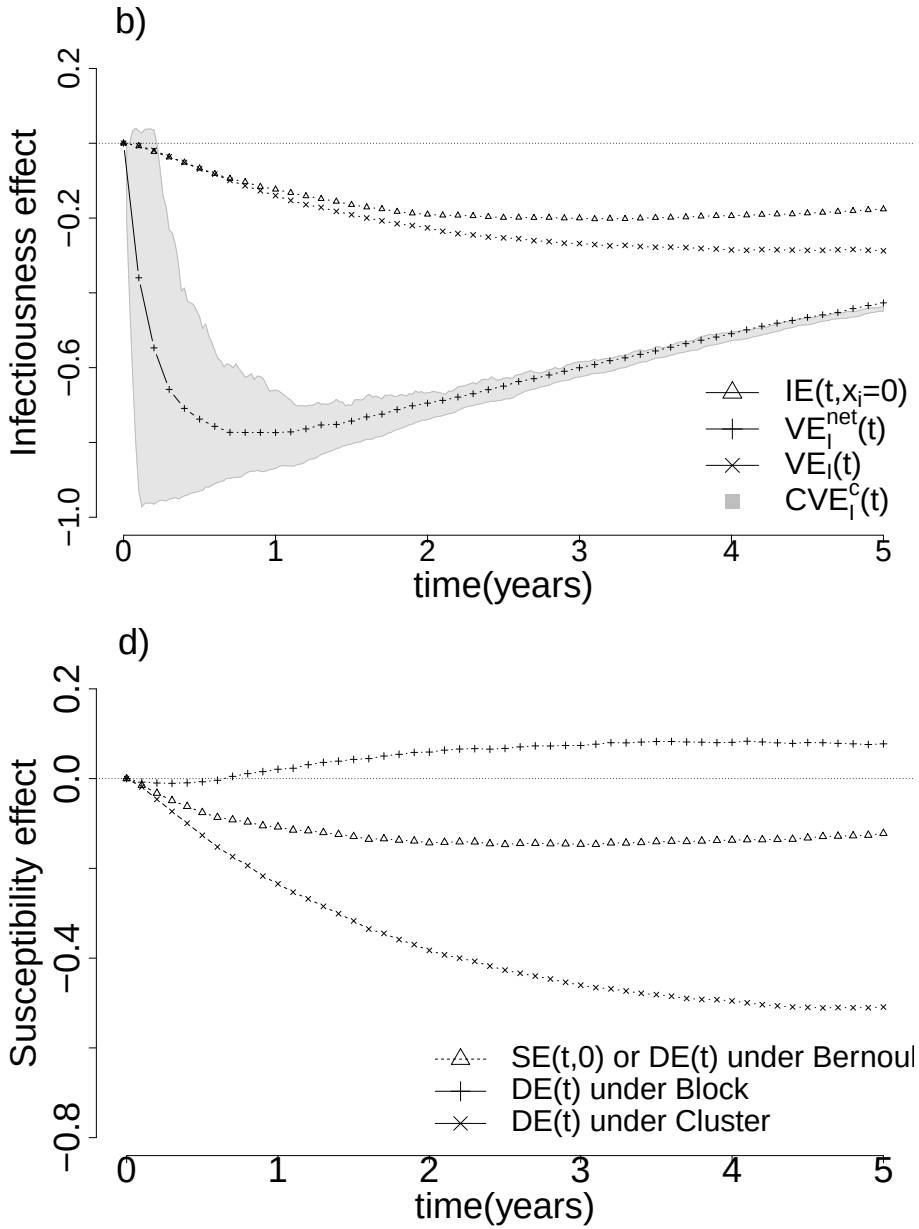
- Exposure-marginalized susceptibility effect $SE(t, 0)$
- Exposure-marginalized infectiousness effect $IE(t, 0)$
- Exposure-marginalized contagion effect $CE(t, 0, 0)$
- Direct effect $DE(t)$
- Indirect effect $IDE(t)$
- ... (others omitted, see details in paper)

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	-0.18	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	-0.21	-0.22
	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	-

[1] VanderWeele et al. Effect partitioning under interference in two-stage randomized vaccine trials. Statistics probability letters, 81(7): 861-869, 2011.
[2] Eck et al. Randomization for the susceptibility effect of an infectious disease intervention. Submitted, 2019.

Simulation: Estimations of natural causal estimands



Summary

- we propose a new framework to articulate causal structure of infectious disease outcomes and treatments, in the case of partnership models.
- A class of fundamental (controlled- and marginalized-) causal estimands for the susceptibility, infectiousness and contagion effect of vaccines are proposed and identified non-parametrically.
- Comprehensive comparisons of our casual estimands are made to other popular estimands in contemporary epidemiology.

Takehome message

- Causal estimands involving interactions with others (contagion and infectiousness effects) require control on others' infection times.
- Valid causal estimands depend on proper randomization scheme and proper control of an comparable distribution of the partner's infection time.
- Causal estimands are functions of observation time, underlying transmission dynamic, cluster size and many other factors.
- Direct comparison of causal estimands from different trails with different observational time and other features may not be meaningful.

Further questions

1. How do we generalize the identification technique to bigger clusters?
2. How do we derive causal estimands, which do not depends on observation time, cluster size, transmission dynamic, or interactions between individuals?
3. How do we calculate causal estimands more efficiently for bigger clusters.

Welcome to the causal inference reading group at 2:00 - 3:30 pm on December 1st, 2020 for further discussions about: "Causal identification of infectious disease intervention effects in a clustered population." Xiaoxuan Cai, Eben Kenah, and Forrest W. Crawford.

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.

xc2577@cumc.columbia.edu
https://xiaoxuan-cai.github.io/

https://arxiv.org/abs/1912.04151

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:

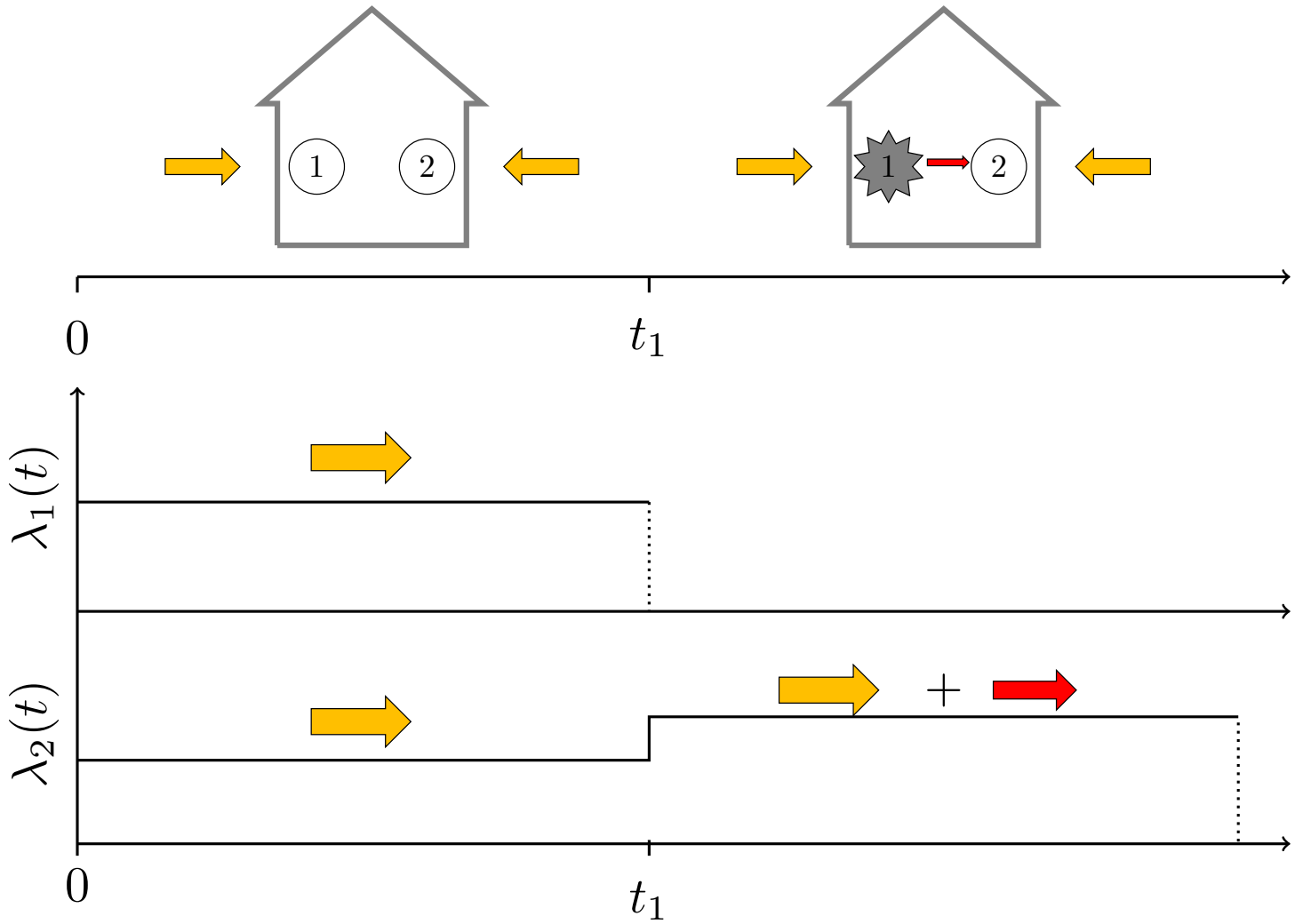
$$\lambda_{0i}(t \mid x_i, \mathsf{I}_i) = \alpha(t) \exp[\beta_1 x_i + \theta_1^T \mathsf{I}_i]$$

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

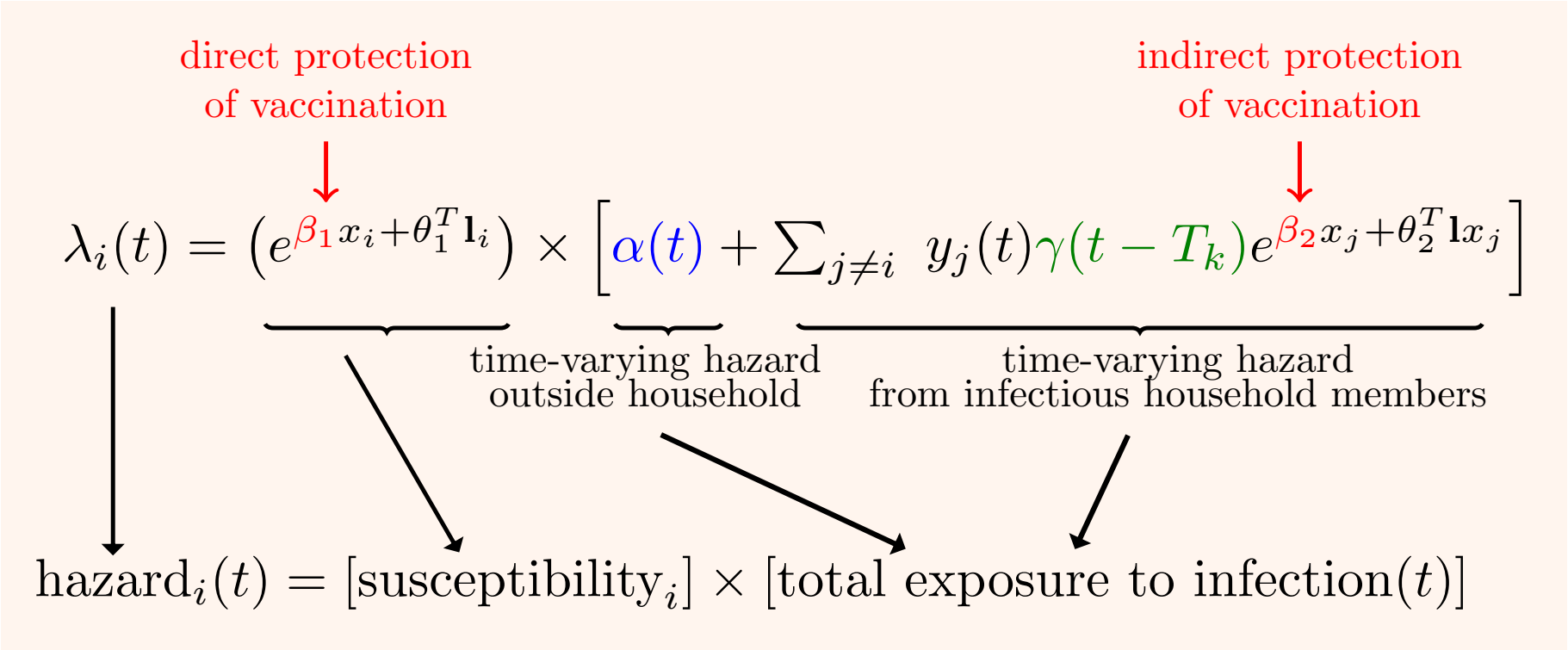
$$\lambda_{ji}(t \mid x_i, \mathsf{I}_i) = \gamma(t - t_j) \exp[\beta_1 x_i + \beta_2 x_j + \theta_1^T \mathsf{I}_i + \theta_2^T \mathsf{I}_j]$$

where $\mathsf{L} = (\mathsf{L}_\mathsf{i}, \mathsf{L}_\mathsf{i})$ measures baseline covariates for the two individuals, including shared covariates for the partnership as a whole.

How do epidemiologists understand infectious disease transmission?



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:

$$HSE^{C}(t, s, x_j, \mathrm{I}) = \frac{\lambda_i(t \mid s, 1, x_j, \mathrm{I})}{\lambda_i(t \mid s, 0, x_j, \mathrm{I})} = e^{\beta_1}$$

- Infectiousness hazard ratio:

$$HIE^{C}(t, s, s', x_i, \mathrm{I}) = \frac{\lambda_i(t|s', x_i, 1, \mathrm{I}) - \lambda_i(t|s, x_i, 1, \mathrm{I})}{\lambda_i(t|s', x_i, 0, \mathrm{I}) - \lambda_i(t|s, x_i, 0, \mathrm{I})} = e^{\beta_2}$$

  where $s' < t < s$.

- Contagion cumulative hazard ratio:

$$HCE^{C}(t; s'', s', \mathrm{I}) = \frac{\int_0^t \left[\lambda_i(u;s',0,0,\mathrm{I}) - \lambda_i(u;s,0,0,\mathrm{I})\right] du}{\int_0^t \left[\lambda_i(u;s'',0,0,\mathrm{I}) - \lambda_i(u;s,0,0,\mathrm{I})\right] du} = \frac{\int_{s'}^{s} \gamma(u) du}{\int_{s''}^{s} \gamma(u) du}$$

  where $s' < s'' < t < s$.

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.	-0.91	-5.03	0.12	-0.14	-0.19	-0.16	-0.20
	Bernoulli	-0.90	-4.58	0.12	-0.14	-0.19	-0.16	-0.20
	Block	-0.88	-4.34	-	-	-	0.06	-
	Cluster	-0.91	-5.22	-	-	-	-0.39	-
Constant hazards without contagion								
2	Obs.	-0.91	-4.62	0.00	-0.18	0.00	-0.18	0.00
	Bernoulli	-0.89	-4.61	0.00	-0.18	0.00	-0.18	0.00
	Block	-0.88	-4.61	-	-	-	-0.18	-
	Cluster.	-0.92	-4.87	-	-	-	-0.18	-
Time-varying hazards								
2	Obs.	-0.88	-4.58	0.12	-0.14	-0.20	-0.21	-0.22
	Bernoulli	-0.93	-4.60	0.12	-0.14	-0.20	-0.21	-0.22
	Block	-0.87	-4.70	-	-	-	0.08	-
	Cluster	-0.92	-4.56	-	-	-	-0.50	-