Estimating the risk of influenza under differing distributions of vaccination among university students

Paul Zivich

Assistant Professor
Department of Epidemiology
Gillings School of Global Public Health
University of North Carolina at Chapel Hill

May 13, 2025

Acknowledgements[1]

This presentation is built from parts of my dissertation

Dissertation Committee: Allison Aiello (chair), Alan Brookhart, Michael Hudgens, James Moody, David Weber

Funding: K01AI177102 (current), T32HD091058 (former), U01CK000185 (eXFLU)

Disclaimer: views (and errors) are mine and not those of NIH or colleagues

pzivich@unc.edu pzivich pausalz@bsky.social

[1] Footnotes are for asides or references, but are fair game for questions

Influenza

Influenza causes substantial morbidity, mortality, economic costs

Influenza prevention among university students

- Elevated risk of infection, low vaccination rates, importance for further transmission[2]
- A key prevention strategy is vaccination

Prior research on university students has focused on

- Direct, or unit-treatment, effect of vaccination
- Focused on vaccination uptake as outcome

Not influenza incidence under large scale changes in vaccination

---

[2] Layde et al. *JID* 1980;142(3):347-352, Bednarczyk et al. *Vaccine* 2015;33(14):1659-1663

Motivation

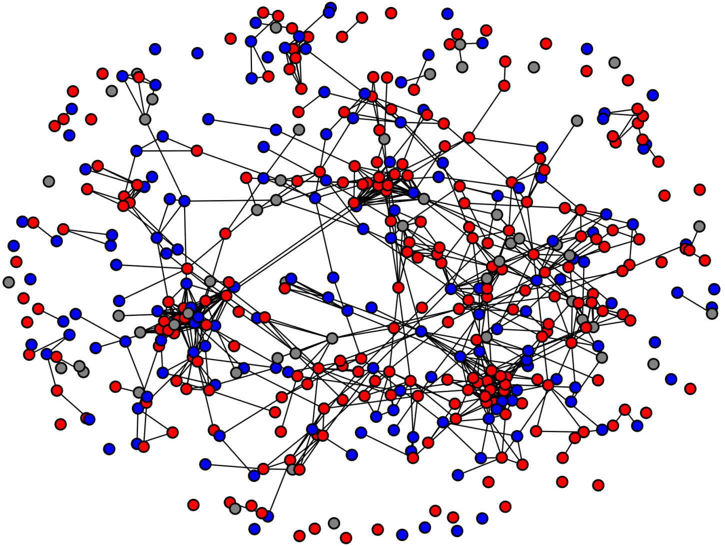
Question: What would the risk of influenza infection be under plans that increase the probability of influenza vaccination uptake among university students at a midwest university from January to April 2013?

Data: eX-FLU cluster randomized trial ($n = 454$)[3]

- Randomized 3-day self-isolation
- 10-weeks of follow-up
- Collected information on vaccination, risk factors for respiratory infections, respiratory infections

[3] Aiello AE et al. 2016 *Epidemics*; 15:38-55

In-Person Contacts Between Students



Challenges

1. Interference & spillover effects
   - Vaccination of one affects another
2. Missing vaccination data
3. Measurement error of contacts

Parameter

Notation

Y_i: observed outcome (i.e., influenza infection) for unit i

A_i: observed action (i.e., vaccination) for unit i

- $\mathcal{A} = \{0, 1\}$ is the support of A
- $\mathbf{A} = (A_1, ..., A_n)$

$Y_i(\mathbf{a}) = Y_i(a_1, ..., a_n) = Y_i(a_i, a_{-i})$: potential outcomes

W_i: vector of covariates for unit i

- $\mathcal{W}$ is the support of W
- $\mathbf{W} = (W_1, ..., W_n)$

$\mathfrak{G}$: $n \times n$ adjacency matrix

- $\mathfrak{G}_{ij} = 1$ if edge between i, j, and 0 otherwise

Parameter of Interest

Given n units in a network $\mathfrak{G}$, parameter is

$$\psi = \frac{1}{n} \sum_{i=1}^{n} E\left[\sum_{\mathbf{a} \in \mathcal{A}} Y_i(\mathbf{a}) \Pr^*(\mathbf{A} = \mathbf{a} \mid \mathbf{W}) \mid \mathbf{W} \right]$$

Mean of Y if $\mathbf{A}$ had been set according to the plan $\Pr^*$, holding the network structure, $\mathfrak{G}$, and $\mathbf{W}$ fixed

Interpretation: Under vaccination plan $\Pr^*$, the incident proportion of influenza would have been ψ for given network

Stochastic Plans

Vaccination Plans

Consider hypothetical interventions that address common reasons for not receiving the influenza vaccine

- **Educational**: correct misconceptions about influenza
- **Non-financial**: ease, availability
- **Financial**: remove monetary costs

Plans

- Based on self-reported reasons
- Shift probability of being vaccinated
- Not vaccinate those with contraindications

Shifts Under Vaccination Plans

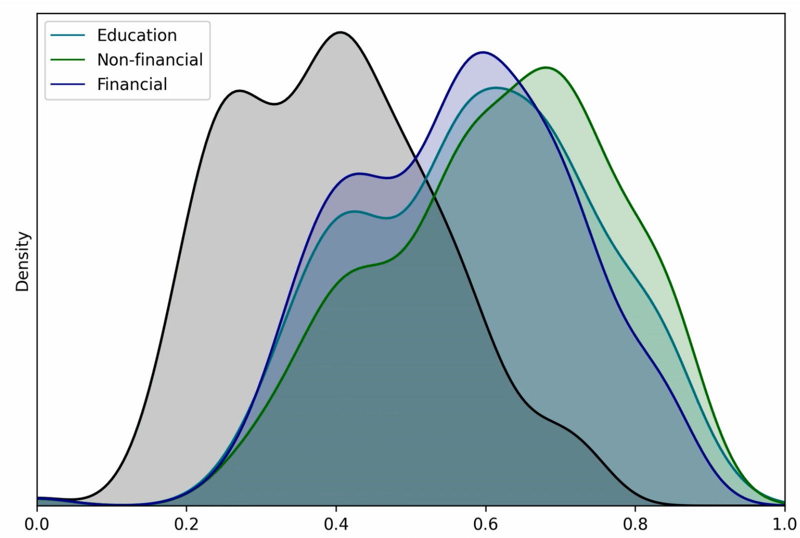
Plan specified in terms of $\Pr^*(A_i = a \mid \mathbf{W}_i)$

Estimate $\rho = \Pr(A_i = 1 \mid \mathbf{W}_i)$ then shift by

$$\rho^* = \begin{cases} \text{expit}\left[\text{logit}(\rho) + \omega\right] & \text{if targeted} \\ \text{expit}\left[\text{logit}(\rho) + \omega/3\right] & \text{otherwise} \end{cases}$$

for $\omega \in [0, 3]$

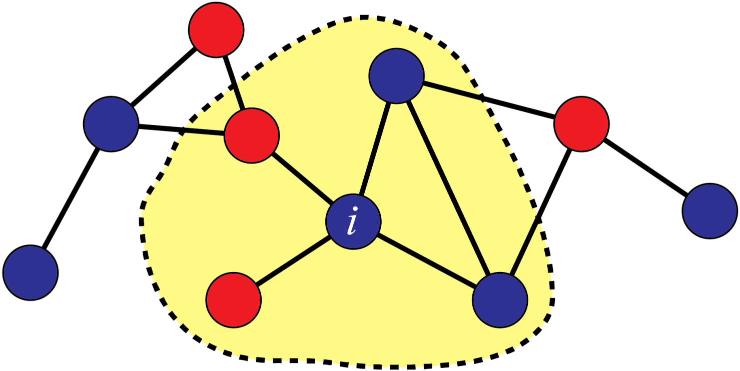
Probability Shifts Under Plans



Identification

Simplifying Interference – Weak Dependence

Learning $Y_i(\mathbf{a})$ is difficult



Covariate Mappings

Rely on a *parametric mapping*

$$Y_i(a_i, a_{-i}) = Y_i(a_i, a_i^s) \neq Y_i(a_i)$$

where a_i^s is a parametric map of i's direct contacts

Here a^s denotes a general covariate mapping of direct contacts

- But need to choose something specific[4]
- Assume parametric mapping is correct

---

[4] These definitions treat all contacts as equivalent but relax via edge weights or multiple edge types

Covariate Mapping Examples

Count

$$X_i^s := \sum_{j \in n} X_j \mathfrak{G}_{ij}$$

Proportion

$$X_i^s := \frac{\sum_{j \in n} X_j \mathfrak{G}_{ij}}{\sum_{j \in n} \mathfrak{G}_{ij}}$$

Threshold

$$X_i^s := I \left\{ \left(\sum_{j \in n} X_j \mathfrak{G}_{ij} \right) >= t \right\}$$

Identification Assumptions

Causal consistency

$$\text{if } a = A_i, a^s = A_i^s \text{ then } Y_i = Y_i(a_i, a_i^s)$$

Exchangeability

$$Y_i(a, a_s) \amalg A_i, A_i^s \mid W_i, W_i^s$$

Positivity

$$\text{if } \Pr^*(A = a, A^s = a^s \mid W, W^s) > 0 \text{ then}$$
$$\Pr(A = a, A^s = a^s \mid W, W^s) > 0$$

Other Systematic Errors

Missing Data

Missing data on vaccination and covariates

Multivariate Imputation with Chained Equations (MICE)

- Modified MICE to include A^s, W^s in models

100 imputed data sets

Measurement Error of Contacts

Self-reported contacts are known to be misreported[5]

Multiple Imputation for Measurement Error (MIME)

- Bayesian procedure with stochastic block model[6]
- Sensitivity and specificity informed by Bluetooth data collected on subset of students

100 networks for each imputed data set

- Summarized 10,000 imputations using nested Rubin's Rule

[5]Mastrandrea et al. *PloS ONE* 2015;10(9):e0136497
[6]Young et al. *Journal of Complex Networks* 2021;8(6)

Estimation

Targeted Maximum Likelihood Estimation (TMLE)

TMLE for network dependent data[7]

1. Fit nuisance model for $Y \mid A, A^s, W, W^s$
2. Fit nuisance model for $A, A^s \mid W, W^s$
3. Fit targeting model
4. Targeted prediction(s)
5. Inference via influence function

[7] van der Laan *Journal of Causal Inference* 2014;2(1):13-74, Sofrygin & van der Laan *Journal of Causal Inference* 2017;5(1):20160003, Zivich et al. *Stats in Med* 2022;41(23):4554-4577

Step 1: Outcome Nuisance Model

Treat observations as if they are independent and fit a model for[8]

$$E[Y \mid A, A^s, W, W^s]$$

Then generate predicted values from this model, $\hat{Y}$, with A, A^s

Here, logistic regression with L_2 penalty

[8] This is the g-computation analog for network-dependent data

Step 2: Action Nuisance Model

For the targeting step, need[9]

$$\frac{\pi_i^*}{\pi_i} = \frac{\Pr^*(A_i = a, A_i^s = a_i^s \mid W_i, W_i^s)}{\Pr(A_i = a, A_i^s = a_i^s \mid W_i, W_i^s)}$$

Factor A, A^s and model as if independent

Use a Monte Carlo procedure to go from

$$\Pr^*(A_i = a \mid W_i, W_i^s) \rightarrow \Pr^*(A_i = a, A_i^s = a^s \mid W_i, W_i^s)$$

Here, logistic regression and Poisson[10]

---

[9]This is the weight for the IPW analog with network-dependent data

[10]A^s was chosen to be count mapping

Step 3: Targeting Step

Fit the following weighted model

$$\text{logit}\left\{\Pr(Y_i = 1)\right\} = \eta + \text{logit}(\hat{Y}_i)$$

where weights are $\frac{\pi_i^*}{\pi_i}$ from the previous step

Step 4: Interest Parameter

Monte Carlo procedure

a. Set actions according to $\Pr^*(A_i = a \mid W_i, W_i^s)$.[11]
b. Compute $\hat{Y}^*$ under A^*, A^{s*}.
c. Update $\hat{Y}^*$ using $\hat{\eta}$, then take mean.
d. Repeat (a.) through (c.) k times.
e. Take average of the k different means as $\hat{\psi}$.

[11] Can simplify by reusing copies from Step 2

Step 5: Inference

Influence function variance estimator with latent dependence[12]

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\frac{\pi_i^*(\hat{\alpha},\hat{\gamma})}{\pi_i(\hat{\alpha},\hat{\gamma})}\left(Y_i - \hat{Y}_i\right)\right] \times \left[\frac{\pi_j^*(\hat{\alpha},\hat{\gamma})}{\pi_j(\hat{\alpha},\hat{\gamma})}\left(Y_j - \hat{Y}_j\right)\right] \times \mathbb{G}_{ij}$$

- where $\mathbb{G}_{ij} = \mathfrak{G}_{ij}$ for $i \neq j$ and $\mathbb{G}_{ij} = 1$ otherwise

[12] Dependence up to second-order contacts. Ogburn EL et al. *JASA* 2024;119(545):597-611

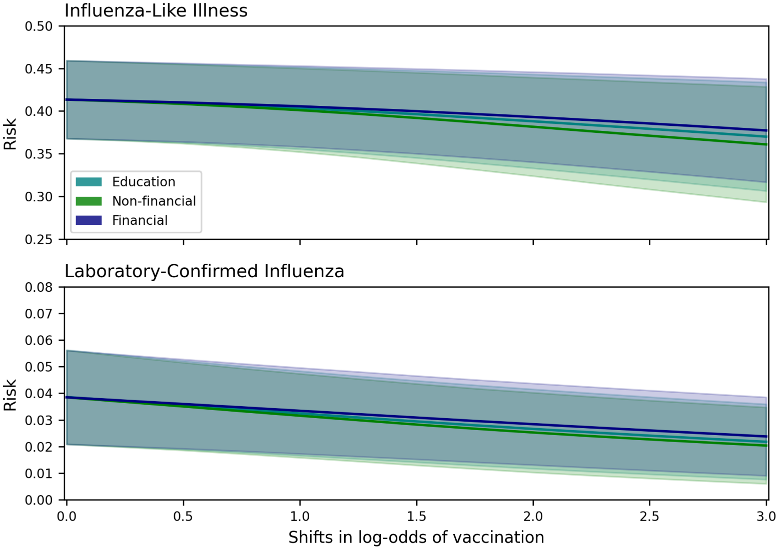
Results

Descriptive Summary

	Overall ($n = 454$)
Influenza-Like Illness	190 (42%)
Laboratory-confirmed	17 (4%)
Vaccination	161 (40%)
Missing	52 (11%)

	Unvaccinated ($n = 241$)
Educational	50 (11%)
Non-financial	93 (20%)
Financial	18 (4%)
Contraindication(s)	2 ($<$1%)

Results


Influenza-Like Illness
Risk
0.50
0.45
0.40
0.35
0.30
0.25
Education
Non-financial
Financial
Laboratory-Confirmed Influenza
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Shifts in log-odds of vaccination

Conclusions

Conclusion

Interference is common in epidemiology

- Modifies how we should view parameters
- Consider what is meaningful for public health
- Need methods (and data) that can accommodate

But there are still many challenges

Practical Challenges

Collection of network data

- Measure large networks
- Reduce measurement error

Modeling

- Weak dependence
- Covariate mappings

Missing data

- Dependence between observations

Partially observing network

Inference doesn't incorporate that $\Pr^*$ is based on estimated probabilities

Future Directions

Longitudinal Network TMLE

- Weak dependence only in interval

Covariate mappings

- Explore flexible specification approaches

Compare against alternative approaches[13]

[13] Tchetgen Tchetgen et al. *JASA*, 2021;116(534):833-844.

Thank You!

Questions?

pzivich@unc.edu pzivich pausalz@bsky.social

Appendix

Identification in Example

Covariates included in W

- gender, race, stress (Perceived Stress Scale-10), optimal hand hygiene, high-risk conditions, sleep quality, alcohol use, trial arm, number of unique contacts

Covariate mappings

- Vaccination: count
- Gender, race, hand hygiene, high-risk, alcohol use, trial arm: count
- stress: variance

Plans

Educational

- Target those: influenza vaccine causes influenza, don't get influenza, didn't know they could get the influenza vaccine

Non-financial

- Target those: never got around to it, did not have transportation, hours available were inconvenient

Financial

- Target those: health plan did not cover, no health insurance

Probability Weights Monte Carlo

Monte Carlo procedure to get π_i^*

a. Create copy of data
b. Set A according to $\Pr^*(A_i = a \mid W_i, W_i^s)$ to get A^*, A^{s*}
c. Repeat (a.) and (b.) for k copies of data
d. Fit models for $\Pr(A_i^* = a, A_i^{s*} = a_i^s \mid W_i, W_i^s)$ using all k copies
e. Predict probability using models from (d.) and *observed* W, W^s, A