Misspecification of Exposure Mapping in Experiments under Interference

Jiayang Nie, Lulude Sun

May 2022

1 Introduction

In this report, we are mainly focusing on the inference tools and results from *Estimating average causal effects under general interference, with application to a social network experiment* by Aronow and Samii. We also make some connections with the factorial experiment paper *Regression-based causal inference with factorial experiments: estimands, model specifications, and design-based properties* by Zhao and Ding. Firstly, we will briefly summarize their methodology and notations. In the second part, we retrieve the data they used and show the summary statistics of the data set. Throughout this report, we are using the field experiment in Aronow and Samii's paper as a showcase example. The paper mainly focuses on HT estimator, and we will justify the use of WLS estimator. Thirdly, we replicate the result with their methodology using weighted least square regression estimate and Horvitz-Thompson Estimator. Next, as the main part of this project, we focus on the consequences of exposure mapping being mis-specified and compare between the HT estimator and WLS estimator by simulation study. The contribution of this project is to explore the consequence of mis-specification, which is a minor gap in Aronow and Samii's original paper. Another contribution is that this project critically assess the use of HT estimator and WLS estimator. We also critically assess the interpretation of the estimates in the context of the field experiment under HT estimate and provide an alternative perspective from factorial regression.

2 Paper Summary

In Zhao and Ding's paper, it provides a way to use linear regression to derive the causal effects for main effects and interaction effects in 2^k experiments. The authors particularly focused on 2^2 factorial experiments as it is more intuitive. Consider two treatments, simultaneously tested on subjects, where we are interested in the

main effect of the first treatment, the main effect of the second treatment and the interaction effect between the two treatments. They can be estimated by running a saturated regression. Zhao and Ding further show that the coefficients from linear regression are unbiased estimators and a Wald type test can be conducted as the covariance matrix $\hat{\Psi}$ asymptotically converge to the actual covariance matrix. Furthermore, $\hat{\Psi}$ under complete randomization is conservative. Therefore, we can construct the causal inference by regression on factorial experiments. We are connecting this idea with Aronow and Samii's paper.

In Aronow and Samii's paper, they provide an elegant way to express causal effects for experimentation under interference. In some literature, this is also called spill-over effect. In this paper, the author primarily focus on the cases when spillover effects exist, and in particular, when the form they may take is well-specified. The critique of using the potential outcome framework for these type of experiments arises from two places. 1. It remains unclear how to interpret the potential outcome as there exists more than treatment and control groups. 2. It is difficult to make unbiased and conservative inference on the difference-in-means estimator because of the unfixed covariance matrix. Specifically, consider the simplest case where there is an indirect effect and a direct effect. Direct effect means unit i receives a treatment, and indirect effect means someone in unit i's social network (like one of unit i's close friends) receives a treatment. Then, instead of the normal experimental design setting where there only exists one difference-in-means estimator, there are five different exposure conditions: 1. direct + no indirect effects, 2. no direct + indirect effects, 3. no direct + no indirect effects, 4. direct and indirect effects. 5. no direct + no indirect effects and the group of unit i is not selected to be the treatment group. For example, if it is exposure condition 4, then both unit i and one of unit i's friends receive the treatment. Let me denote those five exposure conditions as "Direct", "Indirect", "Direct+Indirect", "School", and "No School" respectively. After omitting condition 5, one may notice that this is exactly factorial experiment setting with two factors: Direct and Indirect effects.

2.1 Methodology and Notations

In this paper, the key is that they introduce the notation of exposure condition instead of treatment condition. The difference is that exposure conditions cannot be assigned and the probability of each unit being exposed to certain exposure condition is challenging to find. Note, this probability is fixed and is determined by the randomizing strategy, though is difficult to compute. However, treatment condition can be assigned and causes different exposure condition. In other words, there exists an exposure mapping that maps treatment to exposure conditions. For instance, the researchers can design the randomization such that half of the units will be assigned to treatment. But it is unclear what the chance each unit being under different exposure

condition, which is something jointly determined by the treatment and unit i's social network.

Aronow and Samii also introduce the notation of exposure probability $\pi_i(d_k)$ to be the probability of unit i being under exposure condition k. $\pi_i(d_k)$ can be explicitly computed by iterating through all the combinations of exposure conditions based on units' social network, which is computationally impossible for large dataset. Thus, they proposed a bootstrapping way to sample P different combinations of exposure conditions and taking the average. This estimated exposure probability is denoted by $\pi_i(\hat{d}_k)$. The definitions of exposure probability and exposure condition are important for better understating this report.

They propose an inverse-probability weighted Horowitz and Thompson estimate, as well as weighted least square regression estimates, for estimating causal effects and propose their own covariance matrix estimator to make inference possible. This framework seems to resemble the factorial experiment setting and someone may ask why do they reinvented the wheel. However, the authors' framework are more general: there could be any arbitrary exposure condition, there could be any arbitrary number of treatments, and there could be any arbitrary way to define a network.

2.2 Field Experiment Summary

The field experiment in Aronow and Samii's paper is about a network experiment in American schools. First half of the schools were block randomly selected to participate into the program. Then, students are block randomized by genders, and many other background information and some units will be selected to attend a anti-conflict meeting biweekly. Those who attended anti-conflict meeting can award an orange wristband to fellow students who had anti-conflict behaviors and those orange wristband will be tracked. Note, those students who attended meetings cannot award themselves and can be awarded by another student. In the end, the researchers collect data through a questionnaire to see who is wearing a certified orange band. Therefore, there are five groups: group $d_{001}, d_{011}, d_{101}, d_{111}, d_{000}$. For d_{xyz}, x is 1 when direct exposure exists, y is 1 when indirect exposure exists, and z is 1 when the unit's school is selected to be enrolled in this program. Take d_{101} as an examples, this means direct effect without any indirect effect. Table 1 shows the result we find from their data. We are estimating sample mean for each group here. p is the conditional probability: P(unit i wearing band | unit i is under exposure condition k), and n is the number of units being under exposure condition k. Certainly, this sample proportion is biased to average causal effect because of heteroscedasticity: each unit i can have different probability of being exposed to exposure condition k. The purpose of including this is to validate our data cleaning process and to show that even the biased sample

proportion matches with the paper's result. Also, note that the baseline causal effect d_{000} is 0 in this case. This makes sense because for those schools that were not enrolled in the program, the students certainly cannot wear a certified wristband.

Exposure Condition	p	n
School (d_{001})	0.147	327
Indirect (d_{011})	0.184	277
Direct (d_{101})	0.298	312
Direct+Indirect (d_{111})	0.336	318
No School (d_{000})	0	1200

Table 1: Summary

Table 1 shows the results from the paper and readers can match $\tau(d_{xyz})$ to the exposure conditions we mentioned above for x,y,z in (0, 1). Clearly, the trend is similar for the three types of estimators. Here, $\tau(d_{001}, d_{000})$ is the difference between effect of d_{001} and d_{000} but we find that $d_{000} = 0$. So, in the rest of the paper, we sometimes use $\tau(d_{001})$ as $\tau(d_{001}, d_{000})$.

Social network experiment results: *effects of exposures on probability of wearing a program wristband*

Estimator	**Estimand**	**Estimate**	**S.E.**	**95% CI**
HT	$\tau(d_{001}, d_{000})$	0.057	0.062	$(-0.065, 0.179)$
	$\tau(d_{011}, d_{000})$	0.154	0.029	$(0.097, 0.211)$
	$\tau(d_{101}, d_{000})$	0.305	0.141	$(0.029, 0.581)$
	$\tau(d_{111}, d_{000})$	0.299	0.020	$(0.260, 0.338)$
Hajek	$\tau(d_{001}, d_{000})$	0.058	0.064	$(-0.067, 0.183)$
	$\tau(d_{011}, d_{000})$	0.154	0.037	$(0.081, 0.227)$
	$\tau(d_{101}, d_{000})$	0.292	0.123	$(0.051, 0.533)$
	$\tau(d_{111}, d_{000})$	0.307	0.049	$(0.211, 0.403)$
WLS	$\tau(d_{001}, d_{000})$	0.056	0.066	$(-0.072, 0.186)$
	$\tau(d_{011}, d_{000})$	0.156	0.037	$(0.083, 0.229)$
	$\tau(d_{101}, d_{000})$	0.295	0.124	$(0.050, 0.536)$
	$\tau(d_{111}, d_{000})$	0.306	0.049	$(0.212, 0.404)$

HT = Horvitz–Thompson estimator with conservative variance estimator.
Hajek = Hajek estimator with linearized variance estimator.
WLS = Least squares weighted by exposure probabilities with covariate adjustment for network degree and linearized variance estimator.
S.E. = Estimated standard error; CI = Normal approximation confidence interval.

Figure 1: Paper's Result

3 HT and WLS estimators

3.1 Estimated Exposure Probability

In this paper, the authors use HT estimator and WLS instead of difference-in-means and OLS estimator because the exposure probability is different for different unit and exposure condition. The key difference is the use of exposure probability and thus it is crucial to compute the exposure probability correctly. Figure 2 shows an example of what the exposure probability looks like. pd corresponds to the probability of unit i being exposed to direct effect, pdi corresponds to indirect effect, etc. As can be seen that those probabilities are very random and have little pattern due to the chaotic nature of social network. This is another sign of the existence of heteroscedasticity in the model.

exposure	pd	psch	pdi	pind
Direct+Indirect	0.026875	0.014375	0.248750	0.207500
School	0.261875	0.270000	0.000000	0.000000
No School	0.268125	0.216250	0.000000	0.000000
Direct+Indirect	0.132500	0.125000	0.116250	0.115000
No School	0.057500	0.038125	0.223125	0.165625
No School	0.280625	0.225625	0.000000	0.000000
No School	0.240625	0.267500	0.000000	0.000000
No School	0.006250	0.003750	0.261875	0.236250
Direct+Indirect	0.140625	0.101250	0.133750	0.121875
No School	0.143125	0.123125	0.133125	0.106875
Direct	0.236250	0.261250	0.000000	0.000000

Figure 2: Example of Exposure Probability

3.2 HT Estimator

The WLS and HT estimator are similar in nature, as the relationship between difference-in-means and OLS. It is also easier to justify the assumption with HT estimator first, and then bring it to WLS. The HT estimator is as follows: $\tau(\hat{d}_k) = \sum_{i=1}^{N} I(D_i = d_k)\frac{Y_i}{\pi_i(d_k)}/N$, as opposed to the difference-in-means estimator $\bar{d_k} = \sum_{i=1}^{N} I(D_i = d_k)Y_i/N_{d_k}$ where N_{d_k} is the number of units in exposure condition k. First of all, $\bar{d_k}$ is biased due to the dependency across different exposure conditions. Basically, it is impossible to separate

$I(D_i = d_k)$ and N_{dk} such that they cancel each other out. On the other hand, $\tau(\hat{d_k})$ is unbiased for the inverse-probability weighting. Here, $I(D_i = d_k)$ and $\pi_i(d_k)$ cancels out nicely. The authors spends pages of efforts to propose and prove the conservative confidence interval construction technique. We will not list all the details here.

3.3 WLS Estimator

Similarly, if we believe that the model is $\mathrm{Y} = X\beta + \epsilon$, $\mathrm{E}(\epsilon) = 0$, $cov(\epsilon) = \sigma^2$ for X being five columns of binary values, then the OLS is not the Best Linear Unbiased Estimator anymore. OLS estimates are unbiased, but not BLUE. This is why we need to weight for the inverse of exposure probability. Actually, the estimator from OLS will be the same as the difference in means. The reason why we claim it is biased in the previous paragraph but unbiased now is that it is unbiased under the model assumption but is biased to the average causal effect. The OLS estimate is unbiased only when the multiple linear regression setting makes sense. In mathematics, the WLS regression is $Y \sim 1 + X_{111} + X_{101} + X_{011} + X_{001}$ where the four coefficients excluding the intercept represent the consistent estimators of $d_{111} - d_{000}, d_{101} - d_{000}, d_{011} - d_{000}, d_{001} - d_{000}$. We assume that $\mathrm{Y} = X\beta + \epsilon$, $\mathrm{E}(\epsilon) = 0$, $cov(\epsilon) = \sigma^2\Sigma$. In this case, the estimator under the model assumption is BLUE. Note, the WLS estimator in the paper has some adjustments than this naive WLS estimator. However, authors claim that it will basically be the same as WLS estimator except for extreme forms of clustering. We will be using the simple WLS estimator as it makes no huge difference. Furthermore, we are actually in favor of the WLS estimator than the HT estimator.

The downside of the HT estimator is that it highly relies on the accuracy of estimated exposure probability. As can be seen from the formula: $\tau(\hat{d_k}) = \frac{1}{N}\sum_{i=1}^{N} I(D_i = d_k)\frac{Y_i}{\pi_i(d_k)}$. WLS, however, will not be affected much if the exposure probability is wrongly computed. Algebraically, $\tau_{WLS}(d_k) = \frac{1}{N_k}\sum_{i=1}^{N_k}\frac{Y_i}{p_i}$ where p_i is the probability that unit i is exposed to exposure condition k after normalizing over all the units' probability and N_k is the number of units in exposure condition k. Here, we assume units' indexes are ordered by exposure condition. The key is that the weights here is after normalization. For instance, if weights for HT for unit 1,2,3 is 1/0.01, 1/0.01, and 1/0.02, then the weights for WLS will be 1/0.25, 1/0.25, and 1/0.5. Clearly, the weights in WLS impact the resulting estimator less. Another advantage from WLS is that it is easy to make inference with the standard error of estimated coefficients (Lin 2013). Furthermore, in WLS regression, we can adjust for units' other background information to make the clustering more balanced.

3.4 Factorial Regression Discussion

In the previous subsection, we discuss the assumption of running WLS on five columns of binary variables. Here, we will further discuss the WLS for factorial regression: $Y \sim 1 + A + B + AB$ where A is a binary variable representing having direct effect or not regardless of indirect effect and B is a binary variable representing having indirect effect. The four coefficients (excluding intercept) of the regression are estimate for d_{001}, $d_{101} - d_{001}$, $d_{011} - d_{001}$, $(d_{111} - d_{101}) - (d_{011} - d_{001})$. We omit the d_{000} school-level factor here. The two regression models give ways to constructing the right confidence intervals for different estimands of interest. For instance, if we are interested in finding the four effects separately like the paper, then the regular WLS with five-dummy variables works the best. If we are interested in finding the only the main effects and indirect effects, then it makes more sense to use the factorial regression. It is also easy to make inference for the factorial regression (Zhao and Peng).

3.5 Error in Exposure Probability Estimation

The exposure probability is estimated through bootstrapping. There are two main sources of error:

1. Randomization Error: we run the randomization procedure each time for estimating exposure probability based on our understanding of the paper. There are spaces for misunderstanding and coding errors.

2. Finite Sample Error: the number of all combinations is huge. We are only running a small numbers (1600) of trials to estimate the exposure probability. Because of the limit of computing power, it is impractical to go further. However, 1600 might not be sufficiently large enough and errors may rise from small-sample bias.

If the exposure probability is significantly wrong, then HT may perform worse than WLS.

Another key issue is that the exposure mapping might be wrong. An example of exposure mapping being wrong is there does not exist indirect effect at all. This means that $d_{011} = d_{001}$ and $d_{111} = d_{101}$. If this mis-specification of exposure mapping exists, then the model is no longer 5-dummy regression: $Y \sim 1 + X_{111} + X_{101} + X_{001} + X_{110}$ but 3-dummy regression: $Y \sim 1 + X_{101} + X_{001}$. Anorow and Samii claim that if such mis-specification happens, then the estimated causal effect of the wrong exposure conditions will be distributed to other true exposure conditions. We can roughly have three types of mis-specification: over-complex exposure map, wrongly specified exposure map, and over-simplified model. We already list an example for over-complex exposure map where 1 factor is wrongly specified to be 2. An example of wrongly specified, but equally complex exposure mapping is when units' social network is wrong. An example of over-simplified example is when we only consider direct effect but indirect effect actually exists. We will show later that under mis-specification, HT estimators can sometimes perform way worse than the WLS

estimator.

4 Replication Results

In Table 2, we present the estimate from five-dummy variables Weighted Least Square regression. As can be seen that the result do match that of the paper's result in Figure 1 except for $d_{001} - d_{000}$. For d_{xyz}, x is 1 when direct exposure exists, y is 1 when indirect exposure exists, and z is 1 when the unit's school is selected to be enrolled in this program. Here arises one criticism: the measurement of anti-conflict is solely quantified by the proportion of students wearing a wristband. But clearly for schools that were not selected to participate in this program, they have little chance of wearing. This does not mean those students do not support anti-conflict. The paper includes those students as baseline. The purpose here to include those schools as baseline is perhaps to test the spillover effects among schools, which ends up to be close to zero.

Exposure Condition	Causal Effect	Standard Error
School Exposure ($d_{001} - d_{000}$)	0.158	0.022
Indirect ($d_{011} - d_{000}$))	0.180	0.025
Direct ($d_{101} - d_{000}$))	0.312	0.023
Direct+Indirect ($d_{111} - d_{000}$))	0.343	0.025

Table 2: WLS Replication Result

To interpret Table 2, it means that, in compare to a baseline, the effect of being exposed to isolated school exposure adds 15.8 percent chance to wearing a wristband that represents anti-conflict mindset. This is the school-wide effect of holding anti-conflict events in a school for students who are not directly enrolled in this program or have friends who enrolled in this program. In other words, holding this type of anti-conflict events help all students in that school. Also, the effect of having some friends participated in an anti-conflict event but not themselves adds 18 percent chance of wearing a wristband. This is the social network spillover effect. The effect of being directly enrolled in the program but have no friends enrolled adds 31.2 percent chance of wearing a wristband. Lastly, if one is enrolled in the program and his/her friends are also enrolled in this program, the effect is the greatest: 34.3 percent more chance of wearing a wristband than students in a non event-holder school. Clearly, holding an event about anti-conflict have significant direct and spillover effects on encouraging anti-conflict behaviors. The interesting finding is that indirect and school-level spillover effect seems to be similar in magnitude of school-level spillover effect alone.

4.1 Factorial Regression Interpretation

The paper's focus is on estimating effects of $d_{111} - d_{000}, d_{101} - d_{000}, d_{011} - d_{000}, d_{001} - d_{000}$. However, the critique from us is that d_{000} is close to zero, and it does not seem fair to choose a school that is not

enrolled in the program as a benchmark: students not wearing wristbands may also support anti-conflict courses. Thus, we think a factorial regression is better for examining the main and spillover effects. We run $Y \sim 1 + A + B + AB$ where A is a binary variable representing having direct effect or not regardless of indirect effect and B is a binary variable representing having indirect effect. The four coefficients (excluding intercept) of the regression are estimate for $d_{101} - d_{001}$, $d_{011} - d_{001}$, $(d_{111} - d_{101}) - (d_{011} - d_{001})$. $d_{101} - d_{001}$ is the main effect of being enrolled in the anti-conflict event directly, conditioning on being in a school that enrolled in the anti-conflict program. $d_{011} - d_{001}$ is the spillover effect of having some friends enrolled in anti-conflict events. $(d_{111} - d_{101}) - (d_{011} - d_{001})$ can be interpreted as the average effect of having both direct and indirect effect, excluding direct effect and indirect effect. This estimate is difficult to interpret.

Exposure Factor	Estimate	Standard Error	EHW Standard Error
Main Effect	0.229	0.022	0.037
Spillover Effect	0.096	0.020	0.029
Interaction Effect	-0.066	0.033	0.052

Table 3: Factorial Regression Results

Here, from Table 3, we can see that the main effect of being enrolled into anti-conflict event adds 22.9 percent chance of supporting anti-conflict matters. The spillover effect to peers adds 9.6 percent. This shows a slightly different picture than the previous results. We are simplifying the four effects to be only two effects: main effect and spillover effect.

The last critique we have with the paper's assumption is that the error term ϵ may not follow normal distribution. Actually, since the actual value ranges from 0 to 1, the error term has to be some sort of beta distribution. Therefore, to loose the normality assumption, we also include the EHW robust standard error. Nonetheless, we believe the EHW error here is too conservative as we already adjust for heteroskedasticity.

5 Mis-specification

In section 3.5, we talked about the problem with HT estimator that errors may come from a mis-specified exposure map. Let D_i denote the exposure condition of unit i, which is from the exposure map. Also, $pi_i(d_k)$ is the exposure probability computed from an exposure map. Here, we categories mis-specification into three types: 1. maps that map units into wrong exposure conditions. 2. maps that are finer than the actual exposure map and they map units into the right exposure conditions. 3. maps that are not as fine as the actual exposure map and they map units and they map units into the right exposure conditions. Beside those three types, we also have another two general types: A. the map for estimating exposure probability is wrong but the map for getting exposure condition is correct. B. the map for estimating exposure probability

is wrong and the map for getting exposure condition is wrong in the same way. If the map for estimating exposure probability is wrong and the map for getting exposure condition is wrong in another way, we do not think it is useful to evaluate them as it is too wildly wrong. We will discuss examples of each type. But first, we will look at the cases when exposure map is correctly specified.

5.1 HT and WLS Comparison Under Well-Specification

In this subsection, we report the results from a simulation study. We use the same randomization strategy as the authors used and simulate for 160 times. For each time, we randomly generate a treatment plan, and generate exposure condition accordingly. Here, we assume the exposure map is correct. Everyone is therefore assigned to five exposure groups: Direct, Indirect, Direct+Indirect, No Direct or Indirect. Then, we use rbernoulli to generate an observed value for each unit representing wearing a valid wristband or not. The probability parameter is from the actual paper: probability of wearing a wristband of direct exposure, direct and indirect exposure, direct and indirect exposure and no exposure are set to be 0.299, 0.305, 0.154, 0.057 respectively. Then, we run WLS to obtain the estimate of the five average causal effects. In the end, we obtain 160 samples of HT and WLS estimates assuming known parameters. We compute the mean, standard deviation, and RMSE of the estimates. From Table 4, we can see that HT sometimes performs better than WLS and most of the time WLS out-performs HT. If our exposure probability is more accurate, we believe that HT estimate will be better than WLS. The authors proposed a bound of the bias for HT estimate and as the number of bootstrapping increases to infinity, the bias will converge to zero.

Exposure Condition	Bias (HT)	Sd (HT)	RMSE (HT)
School Exposure ($d_{001} - d_{000}$)	0.000	0.022	0.022
Indirect ($d_{011} - d_{000}$)	-0.049	0.018	0.052
Direct ($d_{101} - d_{000}$)	-0.010	0.022	0.102
Direct+Indirect ($d_{111} - d_{000}$)	0.002	0.041	0.041
Exposure Condition	Bias (WLS)	Sd (WLS)	RMSE (WLS)
School Exposure ($d_{001} - d_{000}$)	0.006	0.028	0.028
Indirect ($d_{011} - d_{000}$)	-0.008	0.024	0.025
Direct ($d_{101} - d_{000}$)	-0.001	0.025	0.025
Direct+Indirect ($d_{111} - d_{000}$)	0.001	0.013	0.013

Table 4: Simulation Study for HT Estimator and WLS Estimator Under Well-Specified Exposure Mapping

5.2 Type 1 and Type A Mis-Specification

An example of Type 1 mis-specification is that when we consider unit i has indirect effect when the one of the top ten nearest peers in his network is treated. However, the actual exposure condition is that indirect effect only exists when the nearest peer in his network is treated. In this case, some of the units

that should be in d_{101} are mapped to d_{111} and some units that should be in d_{001} are mapped to d_{101}. In other words, probability of being under direct exposure and being under school exposure for some units are underestimated. Due to Type A, let us also assume that we only misspecify the map when we estimate the exposure probability but have some means to know the actual exposure mapping when collecting the results. In other words, the only term that is wrong in $\tau(\hat{d}_k) = \frac{1}{N}\sum_{i=1}^{N} I(D_i = d_k)\frac{Y_i}{\pi_i(d_k)}$ is π_i. Note, D_i is mapped by the correct exposure mapping from units' characteristics to an exposure condition. If we use HT estimator, clearly, the effects of those will be over-estimated because of the inverse weighting. From Table 5, we can see that HT estimator under this type of misspecificationi has surprisingly large RMSE. Particularly for d_{001} and d_{101}. The estimate of proportion even passes 1. It is obvious that WLS is way more stable under this misspecification. The reason is that the number of units that are misspecified is small. Thus, WLS estimator is not impacted much. Nonetheless, for those few occurrences, their estimated probability of exposure of d_{101} and d_{001} is under-estimated. It is estimated to be very unlikely to be in those exposure condition. For instance, the probability might be estimated to be 0.001 but the actual probability is 0.1. The weight is then 1000 but those weights appear too often such that the overall estimate is over-estimated. The WLS is not impacted because the weights are in a related scale.

Exposure Condition	Bias (HT)	Sd (HT)	RMSE (HT)
School Exposure ($d_{001} - d_{000}$)	0.208	0.162	0.263
Indirect ($d_{011} - d_{000}$)	-0.134	0.007	0.134
Direct ($d_{101} - d_{000}$)	-0.267	0.010	0.267
Direct+Indirect ($d_{111} - d_{000}$)	1.075	0.399	1.147
Exposure Condition	Bias (WLS)	Sd (WLS)	RMSE (WLS)
School Exposure ($d_{001} - d_{000}$)	-0.013	0.033	0.053
Indirect ($d_{011} - d_{000}$)	0.001	0.055	0.055
Direct ($d_{101} - d_{000}$)	-0.007	0.064	0.065
Direct+Indirect ($d_{111} - d_{000}$)	0.012	0.071	0.071

Table 5: Type 1 Mis-Specification for HT Estimator and WLS Estimator

5.3 Type 2 and Type B Mis-Specification

For Type 2 and Type 3, we assume that it is all Type B errors. In other words, in $\tau(\hat{d}_k) = \frac{1}{N}\sum_{i=1}^{N} I(D_i = d_k)\frac{Y_i}{\pi_i(d_k)}$, π_i and $I(D_i = d_k)$ are wrong in the same way. Note, D_i is mapped by the wrong exposure mapping from units' characteristics to an exposure condition, and π_i is computed from the same wrong exposure mapping. This is because if they are wrong in different ways, as in the above subsection, the HT estimate and WLS is incomparable. Thus, we assume that we misspecify the map both when we estimate the exposure probability and collected the results. For example, what if there does not exist indirect effect at all. This means that $d_{011} = d_{001}$ and $d_{111} = d_{101}$. If this mis-specification of exposure mapping exists, then

the model is no longer 5-dummy regression: $Y \sim 1 + X_{111} + X_{101} + X_{001} + X_{110}$ but 3-dummy regression: $Y \sim 1 + X_{101} + X_{001}$. This is actually the best among the three specification types because the map is correct, but in a sense over-complicate the exposure map. Note that the bias for indirect effect is to compare the estimated one with the school exposure effect, and the bias for indirect + direct exposure is to compare with the direct exposure. When such spill-over effects do not exist, but we still consider them in our framework, we can see that the results are mostly unbiased. In other words, d_{011} is estimated to be the same as d_{001}, and d_{111} is equal to d_{101}, which makes sense. WLS estimator outperforms HT in terms of RMSE and Bias. They have 0 bias. We suspect the reason to be the inaccurate exposure probability estimation.

Exposure Condition	Bias (HT)	Sd (HT)	RMSE (HT)
School Exposure ($d_{001} - d_{000}$)	0.001	0.024	0.024
Indirect ($d_{011} - d_{000}$)	-0.018	0.010	0.021
Direct ($d_{101} - d_{000}$)	-0.092	0.023	0.095
Direct+Indirect ($d_{111} - d_{000}$)	0.003	0.074	0.074
Exposure Condition	Bias (WLS)	Sd (WLS)	RMSE (WLS)
School Exposure ($d_{001} - d_{000}$)	0.000	0.022	0.022
Indirect ($d_{011} - d_{000}$)	0.000	0.015	0.015
Direct ($d_{101} - d_{000}$)	0.000	0.030	0.030
Direct+Indirect ($d_{111} - d_{000}$)	0.000	0.047	0.047

Table 6: Type 2 Mis-Specification for HT Estimator and WLS Estimator

5.4 Type 3 and Type B Mis-Specification

For Type 3 and Type B mis-specification, the example is we assume the exposure mapping only includes $d_{001}, d_{101}, d_{111}, d_{011}$ but there actually exists more than that. What if there exists a second indirect effect? For instance, if units' first peer is treated, then they are 1st indirectly exposed and if units' next top 9 peers are treated, then they are 2nd indirectly exposed. In this case, we are over-simplifying the exposure map. In other words, some spill-over effects are combining into one spill-over effect: 1st and 2nd indirect exposure count toward indirect exposure. Here, it is not reasonable to report the bias for indirect and direct + indirect exposure since they do not exist. From Table 7, We can see that for school exposure, the bias is close to 0 and RMSE is stable for both WLS and HT estimator. However, the estimate of Direct exposure for WLS outperforms HT. We also believe the error source is the estimated exposure probability.

5.5 Complexity of Proposed Exposure Map Trade-off

We can see that if the proposed exposure map is more complex that it should be, then it usually would not cause much error. This can be proved by the fact that Type 3 has larger RMSE than Type 2. However, if the proposed exposure map is too complex, then much error will come from the estimated exposure probability.

Exposure Condition	Bias (HT)	Sd (HT)	RMSE (HT)
School Exposure ($d_{001} - d_{000}$)	0.007	0.060	0.060
Indirect ($d_{011} - d_{000}$)	NA	0.014	NA
Direct ($d_{101} - d_{000}$)	-0.123	0.023	0.126
Direct+Indirect ($d_{111} - d_{000}$)	NA	0.040	NA
Exposure Condition	Bias (WLS)	Sd (WLS)	RMSE (WLS)
School Exposure ($d_{001} - d_{000}$)	0.005	0.022	0.022
Indirect ($d_{011} - d_{000}$)	NA	0.015	NA
Direct ($d_{101} - d_{000}$)	-0.043	0.030	0.030
Direct+Indirect ($d_{111} - d_{000}$)	NA	0.047	NA

Table 7: Type 3 Mis-Specification for HT Estimator and WLS Estimator

Because the more complex exposure map is, the more replications it need to correctly estimate exposure probability for each condition. Therefore, we recommend researchers to find the balance but not go for the most complex or the least complex exposure map.

6 Conclusion

We showed that HT estimator and WLS estimator can be unbiased estimator for causal effects under mis-specification. Furthermore, we empirically proved that there are more sources of error for HT estimator than for WLS estimator. The main sources of error of HT estimator come from mis-estimation of exposure probability but not the misspcification itself. If exposure probability is estimated correctly, we believe the HT estimator will not suffer from misspecification much. However, it is very difficult to compute exposure probability correctly for complex network like social network. Aronow and Samii proved that the HT error is proportional to the smallest exposure probability, with the same number of replications for estimating exposure probability. This means, for complex network, if one of the units is very unlikely to be exposed under one exposure condition, then the error for HT will be large. For example, consider student i that has many friends, and it is very likely for him to be exposed indirectly. However, there still exists a small chance, say 0.02, that student i will be not be exposed to any exposure at all. But in estimating the exposure probability, this may only came up once from 100 trials, and thus the estimated exposure probability is 0.01. Then, the weights for student i will be half as it should be.

We consider the complexity to be composed of exposure map and social network. In our field experiment, it is impossible to lower the complexity of social network. But we are able to lower the complexity of exposure map. Table 5 (Type 2) is a less complex exposure map than Table 4. We see that Table 5 in general has a small bias and RMSE. This empirically proves that the RMSE and bias of estimators move with the complexity of exposure map and social network. One criticism may arise for Table 5 is that it is essentially the ordinary treatment and control framework, and, in this case, the exposure probability is the same as

treatment probability. It is fixed and a diff-in-means can be good enough. Therefore, here we are comparing experiments under no interference with estimators proposed for experiments under interference. It might be unfair. However, we believe this trend follows in general in experiments under interference.

The case when HT estimator may outperform WLS estimator is when exposure probability is estimated correctly. This will happen only if the level of complexity for social network and exposure map is low. Therefore, for complex social network and exposure map, we recommend WLS estimator.

We also show that Type A misspecification is the type to be avoided if we are using HT estimator. Under type A misspecification, HT estimator could be terribly wrong. Some examples are: researchers use a wrong exposure map to map units' characteristics to exposure conditions for estimating exposure probability. However, they collect the correct exposure condition from units. Perhaps, it is because each unit self-report their exposure condition. Therefore, we recommend researchers to always reconfirm with WLS estimator when doing such experiments. If WLS and HT differ in large magnitude, it is a sign of mis-specification.

Under type B misspecification, HT estimator and WLS are comparable in a similar scale. Therefore, we recommend researchers to always use one exposure map for estimating exposure probability and for mapping units to exposure conditions in collecting data. In other words, using a wrong and a correct exposure mapping for is worse than just using the wrong exposure mapping, given that it is not wildly wrong.

We have not gone far enough to test proposed exposure maps that are significantly different from the actual exposure map. We believe the RMSE of both HT and WLS estimators will be large and they will both be unstable.