Stepped Wedge Cluster Randomized Trials

Kenny Chiu

Supervising Faculty: John Petkau

The University of British Columbia
Department of Statistics

STAT 548 Qualifying Paper 2
Oral Presentation

December 9, 2021

Outline

1. Introduction
2. Analysis of stepped wedge cluster randomized trials
3. Investigation of simulation study
4. Extensions to basic model
5. Conclusion

Introduction

Background

Paper by Hussey and Hughes [6] can be viewed as entry point to stepped wedge cluster randomized trials (SW-CRT)

- Provides an overview of motivation, design and analysis of SW-CRTs
- Focuses on technical aspects of practical interest such as power and estimators
- Presents ideas in an accessible and succinct format

Main limitations and weaknesses from our perspective:

1. Limited breadth: discussion is restricted to primarily one SW-CRT setting
2. Minimal depth: technical details are only briefly explained or omitted entirely
3. Writing: unclear which aspects are novel; some typos and/or errors

Objective

Our main goal is to address the limitations of Hussey and Hughes [6]:

1. Address missed technical details, explanations and derivations
2. Clarify their simulation procedure and attempt to replicate their simulation results
3. Discuss extensions to their basic model for different SW-CRT settings

Analysis of stepped wedge cluster randomized trials

Assumed SW-CRT setting

Washington State Community Expedited Partner Treatment (EPT) Trial:

- Hypothesis: EPT public health programs decrease prevalence of chlamydia and incidence of gonorrhea in young women
- Method: Program implemented in 23 local health jurisdictions (LHJ) in 4 waves; primary outcomes were prevalence (incidence) of chlamydia (gonorrhea) in tested women

Primary SW-CRT setting based on EPT trial that Hussey and Hughes [6] work under:

- SW-CRT with $I = 24$ clusters and $T = 5$ measured time points
- Cross-sectional design with $N = 100$ units at each cluster-time

Statistical model

Individual-level model under assumed SW-CRT setting:

$$\begin{aligned} Y_{ijk} &= \mu + \alpha_i + \beta_j + X_{ij}\theta + e_{ijk} \\ &= \mu_{ij} + e_{ijk} \end{aligned}$$

- μ is the mean across clusters and time
- $\alpha_i \sim N(0, \tau^2)$ is a random effect for cluster $i \in \{1, \ldots, I\}$
- β_j is a fixed effect for time point $j \in \{1, \ldots, T-1\}$ ($\beta_T = 0$ for identifiability)
- X_{ij} is a treatment indicator for cluster i at time j (1 denotes intervention)
- θ is the treatment effect of interest
- $e_{ijk} \sim N(0, \sigma^2)$ are i.i.d. noise

Methods for estimating treatment effect θ [6]

1. Within-cluster estimator
   - Consistent if no time effects ($\beta_j = 0$ for all j); biased otherwise [A1]
2. Linear mixed effects model (LMM) via weighted least squares (WLS)
   - Useful if τ^2 and σ^2 known or clusters roughly equal sized; loss of power otherwise due to misspecified weights
   - More efficient than within-cluster estimator if no time effects; note Liao et al. [12] found an error in Hussey and Hughes' relative efficiency [A2]
3. Generalized linear mixed effects model (GLMM)
   - Weights are appropriately weighted even if variance components unknown
   - Link function allows choice of how expected response is modeled
4. Generalized estimating equations (GEE)
   - Consistent even if correlation structure misspecified as long as mean is correctly specified

Power calculation

Hussey and Hughes [6] prescribe using a Wald test to test $H_0 : \theta = 0$

- Power for a test of size α is approximately

$$\Phi \left(\frac{\theta_a}{\sqrt{\mathrm{Var}(\hat{\theta})}} - Z_{1-\frac{\alpha}{2}} \right)$$

where Φ is the cumulative distribution function of a standard normal [A3]

Hussey and Hughes [6] also show that

1. power is maximized when each cluster crosses over at its own time point [A4]
2. delays in treatment effect decreases power [A5]

Investigation of simulation study

Study purpose

- Hussey and Hughes [6] conduct a simulation study to compare powers for testing the treatment effect in LMM, GLMM and GEE
- Their simulation and power calculation procedure is unclear based on their description
- We aim to clarify details of their procedure by attempting to replicate their results

Data simulated based on EPT trial

- $I = 24$, $T = 5$, $\mu = 0.05$, $\tau^2 = 0.000225$
- Risk ratio (RR) chosen for study determines $\theta = \mu(\text{RR} - 1)$

In each of 1000 simulations:

1. Sample cluster effects $\alpha_i \sim N(0, \tau^2)$
2. Shuffle cluster crossover times $t_1, \ldots, t_I$
3. Determine cluster sizes
   - Equal size case: $N_i = 100$ for all i
   - Unequal size case: two-step procedure where

$$p \sim \text{Dirichlet}(1, \ldots, 1)$$

$$\{N_i\}_{i=1}^{I} \sim \text{Multinomial}(99I = 2376, p) + \left[1, \ldots, 1\right]^T$$

4. Sample N_i individuals for cluster i and time j according to Bernoulli(p_{ij}) where

$$p_{ij} = \max(0, \mu + \alpha_i + \mathbf{1}(j \geq t_i)\theta)$$

Model fitting procedure

Compared models (default function arguments used unless otherwise specified):

1. LMM (via lme() from nlme):

$$\mathbb{E}[Y_{ij}|\alpha_i, \beta_j] = \mu + \alpha_i + \beta_j + X_{ij}\theta$$

2. GLMM (via glmmPQL() from MASS) and GEE (via gee() from gee):

$$\mathbb{E}[Y_{ijk}|\alpha_i, \beta_j] = \mu + \alpha_i + \beta_j + X_{ij}\theta$$

   - GEE correlation structure specified to be exchangeable
   - Unclear if Hussey and Hughes use identity or default logit link function

Power calculation procedure

To estimate power:

1. In each simulation, calculate Wald test statistic

$$W = \frac{\hat{\theta}}{\sqrt{\widehat{\mathrm{Var}}(\hat{\theta})}}$$

   and reject if $|W| > z(0.975)$ quantile of standard normal

2. Estimate power = number of rejections / number of non-failing simulations

Hussey and Hughes [6] use two variance estimates:

1. "Standard variance": we interpret as standard error given in function output
2. Jackknife estimate: we use

$$\widehat{\mathrm{Var}}(\hat{\theta}) = \frac{1}{M^2}\sum_{i=1}^{I} N_i^2(\hat{\theta}_i - \hat{\theta}_{\mathsf{JK}})^2 \begin{cases} \hat{\theta}_i = \frac{M\hat{\theta}(\mathbf{y})-(M-N_i)\hat{\theta}(\mathbf{y}_{-i})}{N_i} & \text{cluster pseudo-value} \\ \hat{\theta}_{\mathsf{JK}} = \frac{1}{M}\sum_{i=1}^{I} N_i\hat{\theta}_i & \text{JK estimate of } \theta \end{cases}$$

   where $M = \sum_{i=1}^{I} N_i$ and $\hat{\theta}(\bullet)$ is estimate based on data $\bullet$

Estimated powers using standard variance

Original results from Hussey and Hughes [6]:

	Same cluster sizes			Different cluster sizes		
RR	LMM	GEE	GLMM	LMM	GEE	GLMM
1.0	0.056	0.084	0.076	0.048	0.095	0.069
0.7	0.697	0.719	0.716	0.307	0.703	0.697
0.6	0.907	0.907	0.917	0.487	0.879	0.906
0.5	0.988	0.990	0.992	0.625	0.982	0.986

Our results:

	Same cluster sizes					Different cluster sizes				
	LMM	GEE		GLMM		LMM	GEE[1]		GLMM	
RR		id	logit	id	logit		id	logit	id	logit
1.0	0.050	0.089	0.081	0.066	0.052	0.062	0.11	0.10	0.058	0.053
0.7	0.700	0.736	0.723	0.805	0.711	0.345	0.70	0.68	0.779	0.688
0.6	0.920	0.928	0.920	0.963	0.929	0.536	0.95	0.93	0.951	0.913
0.5	0.981	0.983	0.982	0.994	0.985	0.719[2]	0.99	0.97	0.997	0.985

[1] Estimated over 100 simulations

[2] Estimated over 998 non-failing simulations

Estimated powers using jackknife estimate

Original results from Hussey and Hughes [6]:

	Same cluster sizes			Different cluster sizes		
RR	LMM	GEE	GLMM	LMM	GEE	GLMM
1.0	0.057	0.052	0.053	0.038	0.053	0.049
0.7	0.658	0.644	0.580	0.307	0.577	0.559
0.6	0.884	0.866	0.820	0.503	0.807	0.805
0.5	0.984	0.981	0.948	0.653	0.946	0.942

Our results (estimated over 100 simulations):

	Same cluster sizes			Different cluster sizes		
Risk ratio	LMM	GLMM		LMM	GLMM	
		id	logit		id	logit
1.0	0.06	0.09	0.07	0.02	0.08	0.07
0.7	0.69	0.70	0.70	0.28[3]	0.73	0.69
0.6	0.90	0.95	0.91	0.61	0.93	0.89[3]
0.5	1.00	1.00	1.00	0.66	0.99[3]	0.93[3]

[3] Estimated over 99 non-failing simulations

Extensions to basic model

Model extensions I

1. **Unequal cluster sizes**
   - May not be possible to recruit/maintain equal number of participants in each cluster across time
     - 2017 study: almost half of published trials involved unequal cluster sizes [9]
   - Generally does not require different model but affects cluster-level variances
   - Similar to Hussey and Hughes [6], recent studies [13, 15, 7, 17] involving different contexts found results that suggest unequal sizes lead to decreases in power when size is not accounted for

2. **Delayed treatment effect**
   - Treatment effect may not fully realize over one time period (e.g., [5], [1])
   - One approach to account for delays is to allow treatment indicator $X_{ij} \in [0, 1]$ and take θ as full treatment effect
   - Modeling delays may be avoided by extending time periods to allow treatment effect to fully realize
   - Common consideration in SW-CRT literature but not much recent work

Model extensions II

3. **Non-normal response**
   - Model with normally-distributed responses may not be reasonable (e.g., binary in EPT trial [6], 10-point Likert scale in DECIDE-LVAD trial [2])
   - No standard approach for analyzing non-normal outcomes
   - Power calculation formulas have been proposed for binary and discrete outcomes [18, 19]
   - One gap in SW-CRT literature appears to be the study of non-normal and non-discrete outcomes (e.g., time)

4. **Cohort designs**
   - Same participants may be tracked over multiple time periods (open cohort) or throughout trial (closed cohort) (e.g., INSTTEPP trial [14])
   - Account for repeated measurements by adding individual-level random effects ω_{ik}:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \omega_{ik} + X_{ij}\theta + e_{ijk}$$

   - Recent studies [4, 8, 11] examined various cohort designs with differing objectives

Model extensions III

5. **Hierarchical designs**
   - There may be multiple levels of clustering (e.g., CHANGE trial [10])
   - Account for correlation at different levels by adding random effects accordingly, e.g., cluster-time random effects ω_{ij}:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \omega_{ij} + X_{ij}\theta + e_{ijk}$$

   - General effect of multi-level clustering is the inflation of cluster mean variances at each level [16]

6. **Bayesian approaches**
   - May be desirable to incorporate prior knowledge into model by placing prior distributions on fixed effects and hyperparameters
   - Bayesian SW-CRT models can be fit using Gibbs sampling [3]
   - Recent work found that informative priors reduce calculated sample sizes while bias stays relatively small even if mean is moderately misspecified [20]

Conclusion

Summary

- Paper by Hussey and Hughes [6] is accessible but limited in breadth and depth
- We provided technical details and derivations where absent in the original work
- We obtained similar findings in our simulation study but were unable to replicate exact results
- We discussed extensions to the basic model that were not discussed in the original work

References I

[1] Agius, P. A., Cutts, J. C., Han Oo, W., Thi, A., O'Flaherty, K., Zayar Aung, K., Kyaw Thu, H., Poe Aung, P., Mon Thein, M., Nyi Zaw, N., et al. (2020). Evaluation of the effectiveness of topical repellent distributed by village health volunteer networks against Plasmodium spp. infection in Myanmar: A stepped-wedge cluster randomised trial. *PLOS Medicine*, 17(8):e1003177.

[2] Allen, L. A., McIlvennan, C. K., Thompson, J. S., Dunlay, S. M., LaRue, S. J., Lewis, E. F., Patel, C. B., Blue, L., Fairclough, D. L., Leister, E. C., et al. (2018). Effectiveness of an intervention supporting shared decision making for destination therapy left ventricular assist device: the DECIDE-LVAD randomized clinical trial. *JAMA Internal Medicine*, 178(4):520–529.

[3] Cunanan, K. M., Carlin, B. P., and Peterson, K. A. (2016). A practical Bayesian stepped wedge design for community-based cluster-randomized clinical trials: The British Columbia Telehealth Trial. *Clinical Trials*, 13(6):641–650.

[4] Hooper, R. and Copas, A. (2019). Stepped wedge trials with continuous recruitment require new ways of thinking. *Journal of Clinical Epidemiology*, 116:161–166.

[5] Hughes, J. P., Granston, T. S., and Heagerty, P. J. (2015). Current issues in the design and analysis of stepped wedge trials. *Contemporary Clinical Trials*, 45:55–60.

[6] Hussey, M. A. and Hughes, J. P. (2007). Design and analysis of stepped wedge cluster randomized trials. *Contemporary Clinical Trials*, 28(2):182–191.

[7] Kasza, J., Bowden, R., and Forbes, A. B. (2021). Information content of stepped wedge designs with unequal cluster-period sizes in linear mixed models: Informing incomplete designs. *Statistics in Medicine*, 40(7):1736–1751.

[8] Kasza, J., Hooper, R., Copas, A., and Forbes, A. B. (2020). Sample size and power calculations for open cohort longitudinal cluster randomized trials. *Statistics in Medicine*, 39(13):1871–1883.

References II

[9] Kristunas, C., Morris, T., and Gray, L. (2017). Unequal cluster sizes in stepped-wedge cluster randomised trials: a systematic review. *BMJ Open*, 7(11):e017151.

[10] Lescure, D., Haenen, A., de Greeff, S., Voss, A., Huis, A., and Hulscher, M. (2021). Exploring determinants of hand hygiene compliance in LTCFs: a qualitative study using Flottorps' integrated checklist of determinants of practice. *Antimicrobial Resistance & Infection Control*, 10(1):1–11.

[11] Li, F. (2020). Design and analysis considerations for cohort stepped wedge cluster randomized trials with a decay correlation structure. *Statistics in Medicine*, 39(4):438–455.

[12] Liao, X., Zhou, X., and Spiegelman, D. (2015). A note on "Design and analysis of stepped wedge cluster randomized trials". *Contemporary Clinical Trials*, 45(Pt B):338.

[13] Martin, J. T., Hemming, K., and Girling, A. (2019). The impact of varying cluster size in cross-sectional stepped-wedge cluster randomised trials. *BMC Medical Research Methodology*, 19(1):1–11.

[14] Nease Jr, D. E., Daly, J. M., Dickinson, L. M., Fernald, D. H., Hahn, D. L., Levy, B. T., Michaels, L. C., Simpson, M. J., Westfall, J. M., and Fagnan, L. J. (2018). Impact of a boot camp translation intervention on self-management support in primary care: A report from the INSTTEPP trial and Meta-LARC consortium. *Journal of Patient-Centered Research and Reviews*, 5(4):256.

[15] Ouyang, Y., Karim, M. E., Gustafson, P., Field, T. S., and Wong, H. (2020). Explaining the variation in the attained power of a stepped-wedge trial with unequal cluster sizes. *BMC Medical Research Methodology*, 20(1):1–14.

[16] Teerenstra, S., Taljaard, M., Haenen, A., Huis, A., Atsma, F., Rodwell, L., and Hulscher, M. (2019). Sample size calculation for stepped-wedge cluster-randomized trials with more than two levels of clustering. *Clinical Trials*, 16(3):225–236. PMID: 31018678.

References III

[17] Tian, Z., Preisser, J., Esserman, D., Turner, E., Rathouz, P., and Li, F. (2021). Impact of unequal cluster sizes for GEE analyses of stepped wedge cluster randomized trials with binary outcomes. *medRxiv*.

[18] Wang, J., Cao, J., Zhang, S., and Ahn, C. (2021). Sample size and power analysis for stepped wedge cluster randomised trials with binary outcomes. *Statistical Theory and Related Fields*, 5(2):162–169.

[19] Xia, F., Hughes, J. P., Voldal, E. C., and Heagerty, P. J. (2021). Power and sample size calculation for stepped-wedge designs with discrete outcomes. *Trials*, 22(1):1–10.

[20] Zhan, D., Ouyang, Y., Xu, L., and Wong, H. (2021). Improving efficiency in the stepped-wedge trial design via Bayesian modeling with an informative prior for the time effects. *Clinical Trials*, 18(3):295–302. PMID: 33813906.

Appendix

Within-cluster estimator bias [S9]

Within-cluster estimator [6]:

$$\tilde{\theta} = \frac{1}{I}\sum_{i=1}^{I}\left(\frac{\sum_{j=t_i+1}^{T}\bar{Y}_{ij.}}{T-t_i} - \frac{\sum_{j=1}^{t_i}\bar{Y}_{ij.}}{t_i}\right)$$

If there are non-trivial time effects, bias is

$$\begin{aligned}
\mathrm{bias}(\tilde{\theta},\theta) &= \mathbb{E}[\tilde{\theta}] - \theta \\
&= \frac{1}{I}\sum_{i=1}^{I}\left(\frac{\sum_{j=t_i+1}^{T}(\mu+\alpha_i+\beta_j+\theta)}{T-t_i} - \frac{\sum_{j=1}^{t_i}(\mu+\alpha_i+\beta_j)}{t_i}\right) - \theta \\
&= \frac{1}{I}\sum_{i=1}^{I}\left(\frac{\sum_{j=1}^{T}\beta_j X_{ij}}{T-t_i} - \frac{\sum_{j=1}^{T}\beta_j(1-X_{ij})}{t_i}\right) \\
&= \frac{1}{I}\sum_{j=1}^{T}\beta_j\sum_{i=1}^{I}\frac{t_i X_{ij} - (T-t_i)(1-X_{ij})}{t_i(T-t_i)} \\
&= \sum_{j=1}^{T}\beta_j\sum_{i=1}^{I}\frac{t_i - T(1-X_{ij})}{It_i(T-t_i)}
\end{aligned}$$

Relative efficiency of within-cluster to WLS [S9]

Assume no time effects and let

$$U = \sum_{i=1}^{I} \sum_{j=1}^{T} X_{ij} \qquad V = \sum_{i=1}^{I} \left(\sum_{j=1}^{T} X_{ij} \right)^2$$

Hussey and Hughes [6] efficiency of WLS estimator $\hat{\theta}$ to within-cluster estimator $\tilde{\theta}$:

$$\text{efficiency}(\hat{\theta}, \tilde{\theta}) = \frac{\text{Var}(\tilde{\theta})}{\text{Var}(\hat{\theta})} = \frac{\sum_{i=1}^{I} \left(\frac{1}{t_i} + \frac{1}{T - t_i} \right) \left((ITU - U^2) \frac{\sigma^2}{N} + IT(TU - V)\tau^2 \right)}{I^3 \left(\frac{\sigma^2}{N} + T\tau^2 \right)}$$

Liao et al. [12] (correct) efficiency:

$$\text{efficiency}(\hat{\theta}, \tilde{\theta}) = \frac{\sum_{i=1}^{I} \left(\frac{1}{t_i} + \frac{1}{T - t_i} \right) \left((ITU - U^2) \frac{\sigma^2}{N} + IT(TU - V)\tau^2 \right)}{I^3 T \left(\frac{\sigma^2}{N} + T\tau^2 \right)}$$

Efficiency can be shown to be greater or equal to 1

Outline of derivation of WLS estimator variance

Variance of WLS estimator for θ:

$$\mathrm{Var}(\hat{\theta}) = \frac{IT\frac{\sigma^2}{N}\left(\frac{\sigma^2}{N} + T\tau^2\right)}{(ITU - U^2)\frac{\sigma^2}{N} + IT(UT - V)\tau^2}$$

Define $IT \times 2$ design matrix $\mathbf{X}$ and $IT \times IT$ block diagonal $\mathbf{V}$ with blocks $\mathbf{V}_i$:

$$\mathbf{X} = \begin{bmatrix} 1 & X_{11} \\ \vdots & \vdots \\ 1 & X_{IT} \end{bmatrix} \qquad \mathbf{V}_i = \begin{bmatrix} \tau^2 + \frac{\sigma^2}{N} & \tau^2 & \dots & \tau^2 \\ \tau^2 & \ddots & & \vdots \\ \vdots & & \ddots & \tau^2 \\ \tau^2 & \dots & \tau^2 & \tau^2 + \frac{\sigma^2}{N} \end{bmatrix}$$

Derivation steps:

1. Derive $\mathbf{V}_i^{-1}$ using Sherman-Morrison formula
2. Derive $\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}$ directly from matrix products
3. Derive $\mathrm{Var}(\hat{\theta}) = \begin{bmatrix} 0 & 1 \end{bmatrix} (\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X})^{-1} \begin{bmatrix} 0 & 1 \end{bmatrix}^T$ by inverse of 2×2 matrix

Approximate power [S10]

A Wald test uses the test statistic

$$Z = \frac{\hat{\theta}}{\sqrt{\mathrm{Var}(\hat{\theta})}}$$

which under the null $H_0 : \theta = 0$ is standard normally-distributed

Under a simple alternative $H_a : \theta = \theta_a$, the power of a two-tailed test of size α is

$$\begin{aligned}
&\mathbb{P}\left(Z < -Z_{1-\frac{\alpha}{2}} \,\middle|\, H_a\right) + \mathbb{P}\left(Z > Z_{1-\frac{\alpha}{2}} \,\middle|\, H_a\right) \\
&= \Phi\left(-\frac{\theta_a}{\sqrt{\mathrm{Var}(\hat{\theta})}} - Z_{1-\frac{\alpha}{2}}\right) + \Phi\left(\frac{\theta_a}{\sqrt{\mathrm{Var}(\hat{\theta})}} - Z_{1-\frac{\alpha}{2}}\right) \\
&\approx \begin{cases} \Phi\left(\frac{\theta_a}{\sqrt{\mathrm{Var}(\hat{\theta})}} - Z_{1-\frac{\alpha}{2}}\right) & \text{if } \theta_a \gg 0 \\ \Phi\left(-\frac{\theta_a}{\sqrt{\mathrm{Var}(\hat{\theta})}} - Z_{1-\frac{\alpha}{2}}\right) & \text{if } \theta_a \ll 0 \end{cases}
\end{aligned}$$

Reduced time points on power [S10]

By [A4], larger $\mathrm{Var}(\hat{\theta}) \Rightarrow$ lower power

Variance of within-cluster estimator:

$$\mathrm{Var}(\tilde{\theta}) = \frac{\sigma^2}{NI^2} \sum_{i=1}^{I} \left(\frac{1}{T - t_i} + \frac{1}{t_i} \right)$$

If all but one pair of clusters (WLOG cluster j and I) have a unique crossover time,

$$\begin{aligned}
\sum_{i=1}^{I-1} \left(\frac{1}{T - t_i - 1} + \frac{1}{t_i} \right) + \frac{1}{T - t_j - 1} + \frac{1}{t_j} &= \sum_{i=1}^{I-1} \left(\frac{1}{I - i} + \frac{1}{i} \right) + \frac{1}{I - j} + \frac{1}{j} \\
&> \sum_{i=1}^{I} \left(\frac{1}{I - i + 1} + \frac{1}{i} \right) \\
&= \sum_{i=1}^{I} \left(\frac{1}{T - t_i} + \frac{1}{t_i} \right)
\end{aligned}$$

Time factor is larger in shared crossover case and so variance is larger

Delay in treatment effect on power [S10]

By [A4], larger $\mathrm{Var}(\hat{\theta}) \Rightarrow$ lower power

If $X_{ij} \in [0,1]$ and are known, an unbiased within-cluster estimator for θ is

$$\tilde{\theta} = \left(\sum_{i=1}^{I} \sum_{j=1}^{T} \frac{X_{ij}}{T - t_i} \right)^{-1} \sum_{i=1}^{I} \left(\frac{\sum_{j=t_i+1}^{T} \bar{Y}_{ij.}}{T - t_i} - \frac{\sum_{j=1}^{t_i} \bar{Y}_{ij.}}{t_i} \right)$$

$$\mathrm{Var}(\tilde{\theta}) = \frac{\sigma^2}{N} \left(\sum_{i=1}^{I} \sum_{j=1}^{T} \frac{X_{ij}}{T - t_i} \right)^{-2} \sum_{i=1}^{I} \left(\frac{1}{T - t_i} + \frac{1}{t_i} \right)$$

If $X_{ij} \in (0,1)$ for some cluster i and time j, then

$$\sum_{i=1}^{I} \sum_{j=1}^{T} \frac{X_{ij}}{T - t_i} < \sum_{i=1}^{I} \sum_{j=t_i+1}^{T} \frac{1}{T - t_i} = I$$

$$\Rightarrow \quad \frac{\sigma^2}{N} \left(\sum_{i=1}^{I} \sum_{j=1}^{T} \frac{X_{ij}}{T - t_i} \right)^{-2} \sum_{i=1}^{I} \left(\frac{1}{T - t_i} + \frac{1}{t_i} \right) > \frac{\sigma^2}{N I^2} \sum_{i=1}^{I} \left(\frac{1}{T - t_i} + \frac{1}{t_i} \right)$$

RHS is variance of estimator with no delays and so delay increases variance