Estimating Treatment Effects in Cluster Randomized Trials by Calibrating Covariate Imbalances between Clusters

Zhenke Wu, Constantine Frangakis,
Thomas Louis, Daniel Scharfstein

Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health

27 August 2014

Individualizing Health

Source: http://www.diabetesdaily.com/voices/2014/07/why-one-size-fits-all-doesnt-work-in-diabetes

Evaluation of individualized intervention

1. **Scientific question:** To what extent has the individualized rule improved health outcomes for the entire population? (Policy makers may care more than clinicians)
2. **Statistical question:** How to estimate the overall effect *consistently* and *efficiently*?

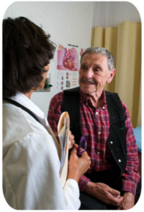
Wu, Frangakis, Louis, Scharfstein (2014). Estimating Treatment Effects in Cluster Randomized Trials by Calibrating Covariate Imbalances between Clusters. *Biometrics*. doi: 10.1111/biom.12214.

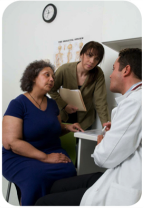
R package: http://github.com/zhenkewu/mpcr

Example: Guided Care study

Background: specially trained nurses to help deliver patient-centered care

Online course and certificate for specially trained nurses





Study website: http://www.guidedcare.org/
Nurse training courses: https://www.ijhn-education.org/content/guided-care-nursing

How data is collected?

Matched-pair cluster randomized (MPCR) design–rationale

1. Sometimes, investigators are only able to intervene on clusters of individuals, e.g., a nurse for each clinical practice

1. Cornfield J (1978)
2. Gail et al. (1992)
3. Moulton L (2004)
4. Imai K, King G, and Nall C (2009)

How data is collected?

Matched-pair cluster randomized (MPCR) design–rationale

1. Sometimes, investigators are only able to intervene on clusters of individuals, e.g., a nurse for each clinical practice
2. To recoup the resulting efficiency loss[1], some studies pair similar clusters and randomize treatments within pairs[2,3]

1. Cornfield J (1978)
2. Gail et al. (1992)
3. Moulton L (2004)
4. Imai K, King G, and Nall C (2009)

How data is collected?

Matched-pair cluster randomized (MPCR) design–rationale

1. Sometimes, investigators are only able to intervene on clusters of individuals, e.g., a nurse for each clinical practice
2. To recoup the resulting efficiency loss[1], some studies pair similar clusters and randomize treatments within pairs[2,3]
3. The use of pre-treatment variables that affect the outcome can improve estimation efficiency[4]

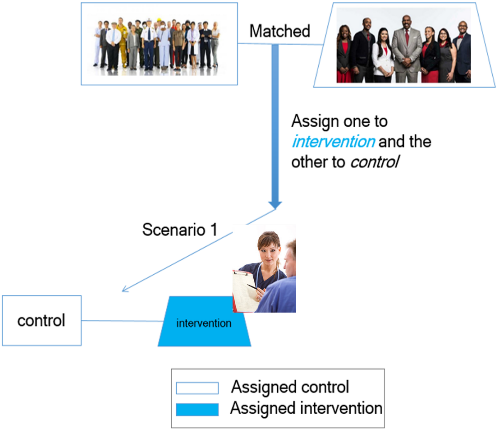
1. Cornfield J (1978)
2. Gail et al. (1992)
3. Moulton L (2004)
4. Imai K, King G, and Nall C (2009)

Matched-pair cluster randomized (MPCR) design

One pair

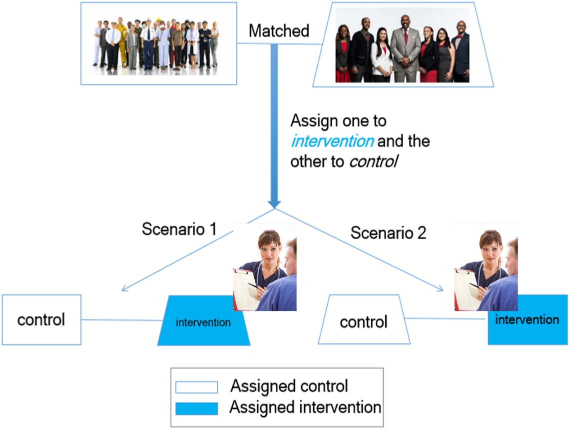
Matched-pair cluster randomized (MPCR) design

One pair



Matched-pair cluster randomized (MPCR) design

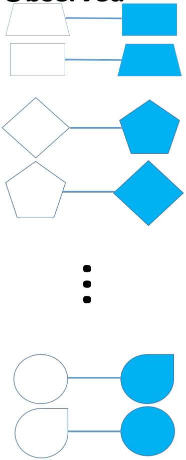
One pair



MPCR design

Example: Guided Care study[5]

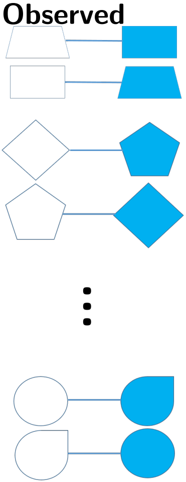
Observed



5. Boult C. et al. (2013)

MPCR design

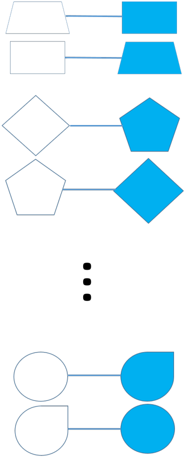
Example: Guided Care study[5]



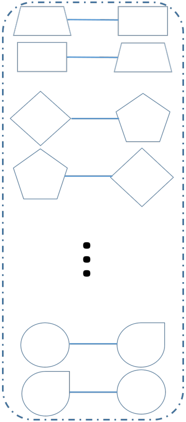
- Intervention: assignment of specially trained nurses to coordinate patient-centered care
- 14 teams of clinical practices matched into 7 pairs
- Covariates: hierarchical condition category (hcc), age, race, gender, education, livesalone, etc.
- Primary outcome: physical component summary in Short-Form 36 (SF-36) Version 2

5. Boult C. et al. (2013)

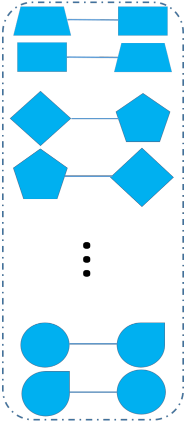
Observed



If all are assigned control



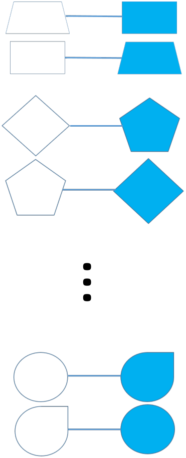
if all are assigned intervention



MPCR design

Goal

Observed

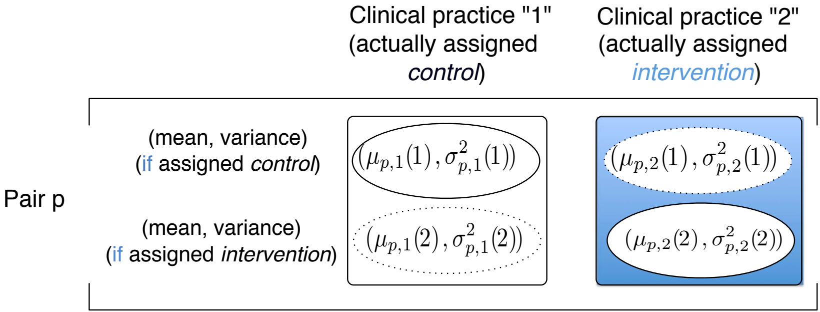


Goal: To estimate the average outcome if all clusters in all pairs are assigned control (1) versus if all clusters in all pairs are assigned to intervention (2):

$$\delta^{\text{effect}} = \mu(1) - \mu(2)$$

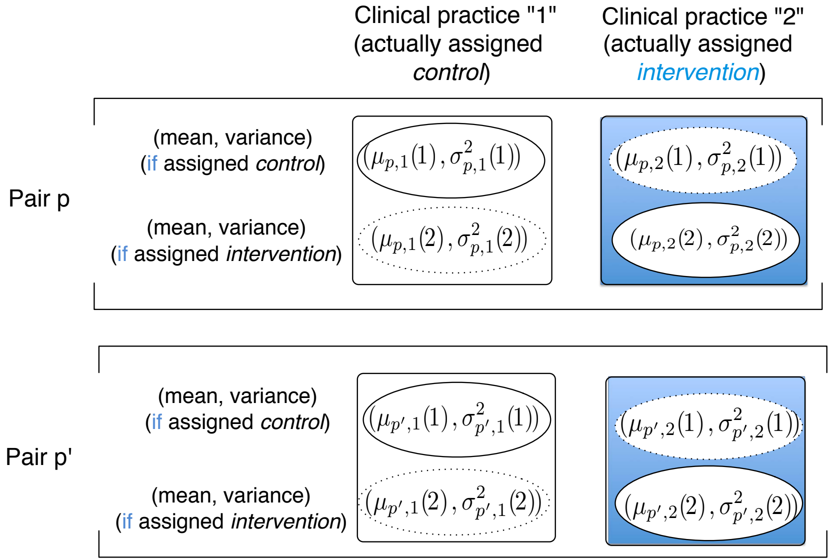
Understanding the observed data from MPCR design

Type 1



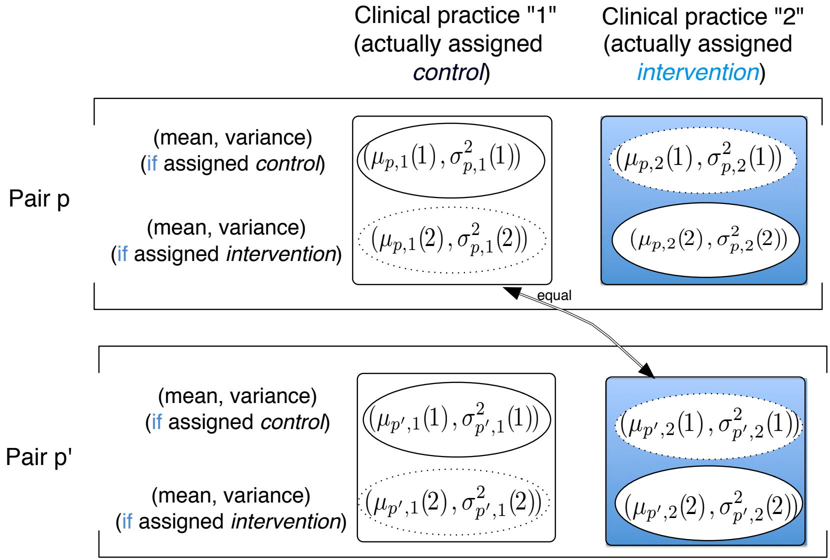
Understanding the observed data from MPCR design

Type 1 and Type 2



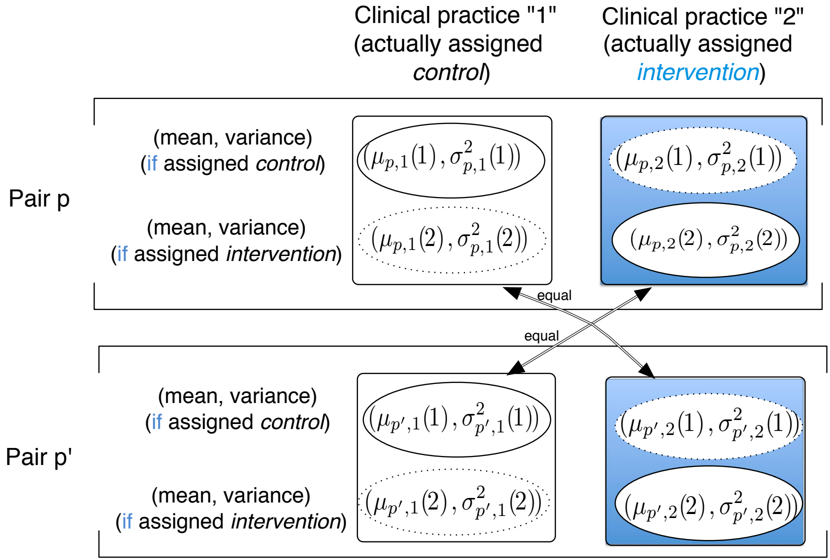
Understanding the observed data from MPCR design

Two types share the same characteristics

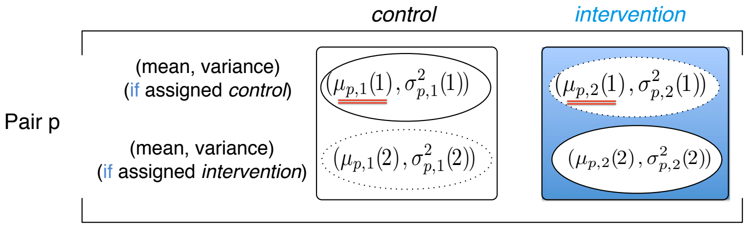


Understanding the observed data from MPCR design

Each type is sampled with probability $\frac{1}{2}$ (design-based)



The right target

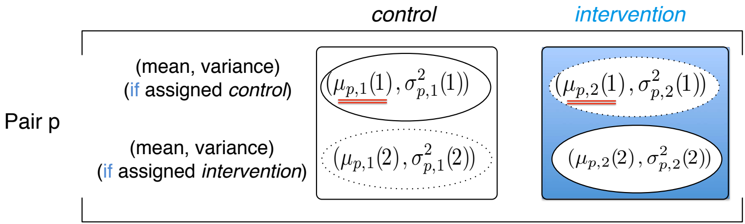


- If all patients are assigned with intervention t,

$$\mu_p(t) = \mu_{p,1}(t)\pi_{p,1} + \mu_{p,2}(t)\pi_{p,2},$$

where $\pi_{p,1}$ is the fraction of patients served by the first clinic; $\pi_{p,2} = 1 - \pi_{p,1}$.

The right target

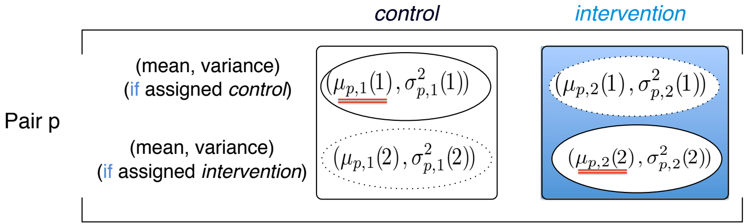


- If all patients are assigned with intervention t,

$$\mu_p(t) = \mu_{p,1}(t)\pi_{p,1} + \mu_{p,2}(t)\pi_{p,2},$$

  where $\pi_{p,1}$ is the fraction of patients served by the first clinic; $\pi_{p,2} = 1 - \pi_{p,1}$.
- Averaging over a population of pairs, $\mu(1) = \mathbb{E}\{\mu_p(1)\}$, $\mu(2) = \mathbb{E}\{\mu_p(2)\}$, $\boxed{\delta^{\text{effect}} = \mu(1) - \mu(2)}$.

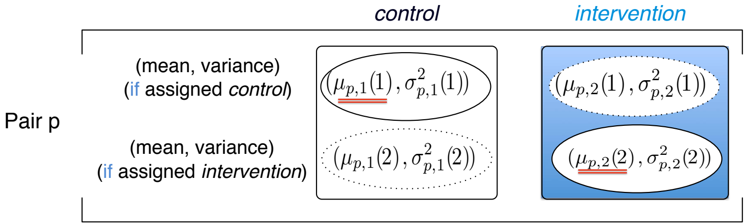
Directly estimable contrasts



- Direct difference between observed means

$$\hat{\delta}_p^{\text{crude}} = \hat{\mu}_{p,1}(1) - \hat{\mu}_{p,2}(2),$$

Directly estimable contrasts



- Direct difference between observed means

$$\hat{\delta}_p^{\text{crude}} = \hat{\mu}_{p,1}(1) - \hat{\mu}_{p,2}(2),$$

with $[\hat{\delta}_p^{\text{crude}} \mid \delta_p^{\text{crude}}, v_p^{2,\text{crude}}]$ approximately normal

Methods for effect estimation under MPCR design

First-level only

Only based on the following equality

$$\underline{\mathbb{E}\left(\delta_p^{\text{crude}}\right) = \delta^{\text{effect}}},$$

without assumptions on $[\delta_p^{\text{crude}}, v_p^{2,\text{crude}}]$.

Methods for effect estimation under MPCR design

First-level only

Only based on the following equality

$$\underline{\mathbb{E}\left(\delta_p^{\text{crude}}\right) = \delta^{\text{effect}}},$$

without assumptions on $[\delta_p^{\text{crude}}, v_p^{2,\text{crude}}]$.

Example: Average of $\hat{\delta}_p^{\text{crude}}$ or other weighted extensions[4]

Methods for effect estimation under MPCR design

With a hierarchical second-level (meta-analysis)

- Directly models observed outcomes, using two-level model[6]

$$\hat{\delta}_p^{\text{crude}} \mid \delta_p^{\text{crude}}, v_p^{2,\text{crude}} \quad \sim \quad \text{Normal}\left(\delta_p^{\text{crude}}, v_p^{2,\text{crude}}\right),$$

6. Thompson et al., (1997)

Methods for effect estimation under MPCR design

With a hierarchical second-level (meta-analysis)

- Directly models observed outcomes, using two-level model[6]

$$
\begin{aligned}
\hat{\delta}_p^{\text{crude}} \mid \delta_p^{\text{crude}}, v_p^{2,\text{crude}} &\sim \text{Normal}\left(\delta_p^{\text{crude}}, v_p^{2,\text{crude}}\right), \\
\delta_p^{\text{crude}} \mid \tau^2 &\sim \text{Normal}\left(\delta^{\text{effect}}, \tau^2\right)
\end{aligned}
$$

6. Thompson et al., (1997)

Methods for effect estimation under MPCR design

With a hierarchical second-level (meta-analysis)

- Directly models observed outcomes, using two-level model[6]

$$\begin{aligned} \hat{\delta}_p^{\text{crude}} \mid \delta_p^{\text{crude}}, v_p^{2,\text{crude}} &\sim \text{Normal}\left(\delta_p^{\text{crude}}, v_p^{2,\text{crude}}\right), \\ \delta_p^{\text{crude}} \mid \tau^2 &\sim \text{Normal}\left(\delta^{\text{effect}}, \tau^2\right) \end{aligned}$$

- **Question:** an implicit assumption in the second level ?

6. Thompson et al., (1997)

Methods for effect estimation under MPCR design

With a hierarchical second-level (meta-analysis)

- Directly models observed outcomes, using two-level model[6]

$$\begin{aligned} \hat{\delta}_p^{\text{crude}} \mid \delta_p^{\text{crude}}, v_p^{2,\text{crude}} &\sim \text{Normal}\left(\delta_p^{\text{crude}}, v_p^{2,\text{crude}}\right), \\ \delta_p^{\text{crude}} \mid \tau^2 &\sim \text{Normal}\left(\delta^{\text{effect}}, \tau^2\right) \end{aligned}$$

- **Question:** an implicit assumption in the second level ?

$$\delta_p^{\text{crude}} \perp\!\!\!\perp v_p^{2,\text{crude}} \mid \tau^2$$

6. Thompson et al., (1997)

Methods for effect estimation under MPCR design

With a hierarchical second-level (meta-analysis)

- Directly models observed outcomes, using two-level model[6]

$$\hat{\delta}_p^{\text{crude}} \mid \delta_p^{\text{crude}}, v_p^{2,\text{crude}} \sim \text{Normal}\left(\delta_p^{\text{crude}}, v_p^{2,\text{crude}}\right),$$
$$\delta_p^{\text{crude}} \mid \tau^2 \sim \text{Normal}\left(\delta^{\text{effect}}, \tau^2\right)$$

- **Question:** an implicit assumption in the second level ?

$$\delta_p^{\text{crude}} \perp\!\!\!\perp v_p^{2,\text{crude}} \mid \tau^2$$

- Can lead to inconsistent effect estimator if not true!

Example of inconsistent estimation

6. Thompson et al., (1997)

Another practical problem: covariate imbalance despite matching

Data from the Guided Care study

	pair						
	1	2	3	4	5	6	7
age at interview[a]	0.3	-0.3	0.1	0.6	0.0	0.1	-0.1
Chronic Illness Burden[a]	0.5	-0.6	0.0	0.0	-1.1	0.1	0.6
SF36 Mental[a]	-0.3	0.1	0.3	0.2	0.3	-0.6	-0.5
SF36 Physical[a]	-0.1	-0.4	0.1	0.5	0.4	-0.6	-0.3

Standardized differences of several continuous covariates between two clusters within each of 7 pairs.

Proposed method: covariate calibration

- **Bias consideration:** If a hierarchical second level is used, to make the following more plausible:

$$\delta_p^{\text{crude}} \perp\!\!\!\perp v_p^{2,\text{crude}} \mid X, \tau^2$$

Proposed method: covariate calibration

- **Bias consideration:** If a hierarchical second level is used, to make the following more plausible:

$$\delta_p^{\text{crude}} \perp\!\!\!\perp v_p^{2,\text{crude}} \mid X, \tau^2$$

- **Efficiency consideration:** To decrease residual variance by conditional on important covariates that affect outcomes

Methods to handle covariate imbalance

Existing approaches

1. Interpretation of treatment effect conditional on covariates[6]
2. Normal assumption on individual level: does not necessarily hold; interpretation of treatment effect conditional on cluster-specific random effects, thus treatment effect require a model to be estimable[7,8]

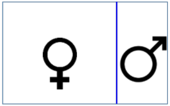
7. Feng et al. (2001)
8. Hill J. and Scott M. (2009)

Covariate-calibrated estimation

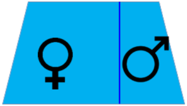
1 Combine covariate distribution, and 2 re-weight outcome regression

1. Stratify the average outcome by covariate

75% Female, n=100



85% Female, n=200

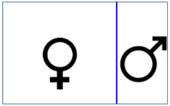


Combined covariate distribution P(x=F)

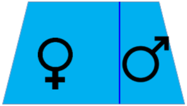
$$75\%\frac{1}{3} + 85\%\frac{2}{3} = 82\%$$

1. Stratify the average outcome by covariate

75% Female, n=100



85% Female, n=200



Combined covariate distribution P(x=F)

$$75\%\frac{1}{3} + 85\%\frac{2}{3} = 82\%$$

2. Re-calibrating the stratified means with respect to the covariate distribution of the two clusters combined, for example, for the control arm $t = 1$,

$$\begin{aligned} \mu^{\text{calibr}}_{p,c=1} &= \int_{x} \mu_{p,c=1}(x; t = 1)\mathrm{d}G_p(x), \\ &= 82\% \cdot \mu_{p,c=1}(x = F; t = 1) \\ &\quad +18\% \cdot \mu_{p,c=1}(x = M; t = 1). \end{aligned}$$

Uncalibrated vs calibrated analysis

Reduced variances

	pair p						
	1	2	3	4	5	6	7
sample size							
$n_{p,c=1}$	17	16	42	23	52	23	28
$n_{p,c=2}$	38	44	43	33	42	31	43
outcome							
	uncalibrated on covariates						
$\hat{\mu}_{p,1}(1)$	36.4	36.5	39.6	39.1	39.7	33.8	39.6
$\hat{\mu}_{p,2}(2)$	37.3	36.6	39.3	35.3	35.2	36.4	40.9
$\hat{\delta}_p^{\text{crude}}$	-0.8	-0.1	0.3	3.8	4.5	-2.6	-1.3
$\left(v_p^{\text{crude}}\right)^{1/2}$	2.7	2.6	2.0	2.7	2.1	2.6	2.2
	calibrated on covariates						
$^*\hat{\mu}_{p,1}^{\text{calibr}}$	37.6	38.8	39.5	38.0	38.7	35.5	40.9
$^*\hat{\mu}_{p,2}^{\text{calibr}}$	36.7	35.8	39.4	36.0	36.4	35.1	40.0
$\hat{\delta}_p^{\text{calibr}}$	0.9	3.0	0.1	1.9	2.3	0.5	0.8
$\dagger\left(v_p^{\text{calibr}}\right)^{1/2}$	2.1	2.4	1.5	2.0	1.7	2.2	1.7

Analysis of Guided Care data

Table: Results from MLE, profile MLE, Bayes estimates and permutation test in the Guided Care study. The outcome is the physical component summary of the Short Form 36 (SF36).

		$\hat{\delta}^{\text{effect}}$	95% C.I.	s.e.($\hat{\delta}^{\text{effect}}$)	$\widehat{\text{var}}(\delta_p^*)$	p-value (two-sided)
uncalibrated on covariates	*1st level*					
	MLE	0.5	$(-1.4, 2.5)$	1.0	–	0.59
	permutation	–	–	–	–	0.61
	1st+2nd level					
	MLE	0.6	$(-1.2, 2.5)$	0.9	0.7	0.50
	pMLE	0.6	$(-1.5, 2.7)$	–	0.7	–
	Bayes	0.6	$(-1.7, 3.0)$	1.2	4.3	0.60
	permutation	–	–	–	–	0.60
calibrated on covariates	*1st level*					
	MLE	1.4	$(0.5, 2.2)$	0.4	–	<0.01
	permutation	–	–	–	–	0.02
	1st+2nd level					
	MLE	1.2	$(-0.2, 2.6)$	0.7	0.0	0.08
	pMLE	1.2	$(-0.2, 2.6)$	–	0.0	–
	Bayes	1.3	$(-0.4, 2.9)$	0.9	1.5	0.13
	permutation	–	–	–	–	0.02

*: represents δ_p^{crude} for the uncalibrated approach and δ_p^{calibr} for the calibrated approach.

Summary

Goal: to evaluate individualized interventions for a population

Summary

Goal: to evaluate individualized interventions for a population
Data: from matched-pair cluster randomized (MPCR) design.

Summary

Goal: to evaluate individualized interventions for a population
Data: from matched-pair cluster randomized (MPCR) design.
Statistical contributions:

- Existing approaches only model the observed data (e.g., meta-analysis). We connect them with potential outcome framework and reveal implicit assumptions

Summary

Goal: to evaluate individualized interventions for a population
Data: from matched-pair cluster randomized (MPCR) design.
Statistical contributions:

- Existing approaches only model the observed data (e.g., meta-analysis). We connect them with potential outcome framework and reveal implicit assumptions
- Covariate-calibration is necessary if 2nd-level checking reveals substantial dependence

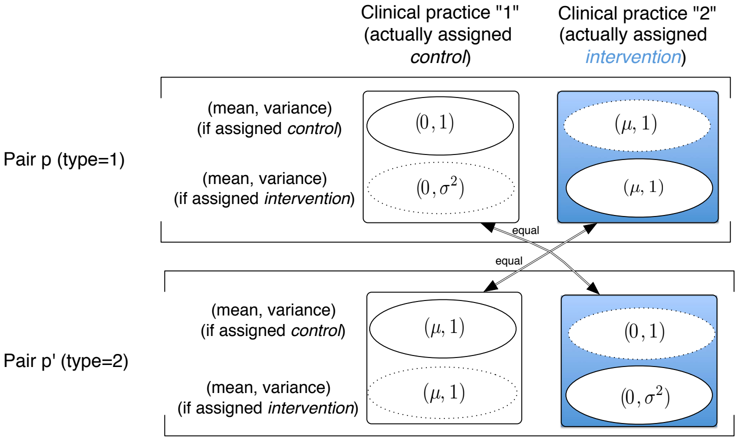
Summary

Goal: to evaluate individualized interventions for a population
Data: from matched-pair cluster randomized (MPCR) design.
Statistical contributions:

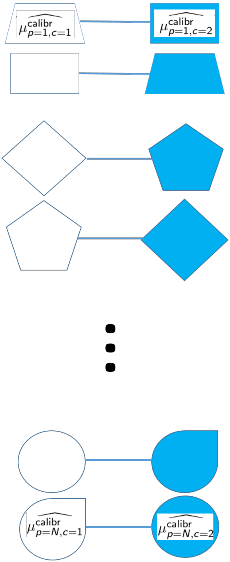
- Existing approaches only model the observed data (e.g., meta-analysis). We connect them with potential outcome framework and reveal implicit assumptions
- Covariate-calibration is necessary if 2nd-level checking reveals substantial dependence
- Covariate-calibration improves estimation efficiency

Thank you!

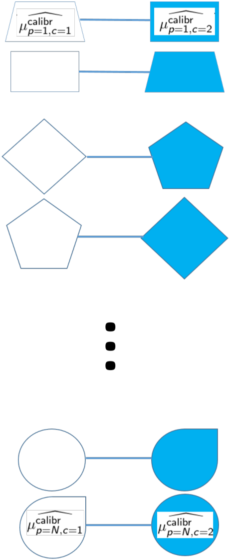
An example of inconsistency of meta-analytic "MLE"



Matched-pair cluster randomized design



Matched-pair cluster randomized design

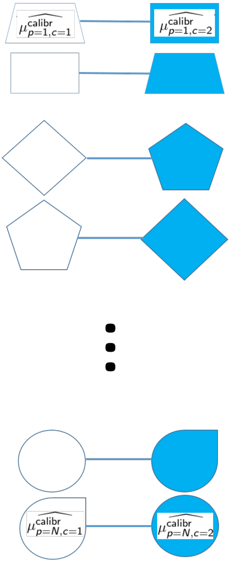


<u>level 1':</u>

$$\begin{bmatrix}\hat{\delta}_1^{\text{calibr}}\\ \vdots \\ \hat{\delta}_N^{\text{calibr}}\end{bmatrix} \Bigg| \begin{bmatrix}\delta_1^{\text{calibr}}\\ \vdots \\ \delta_N^{\text{calibr}}\end{bmatrix}, \theta, \Sigma_{\hat{\delta}^{\text{calibr}}}$$

$$\sim \mathit{Normal}\left\{\begin{bmatrix}\delta_1^{\text{calibr}}\\ \vdots \\ \delta_N^{\text{calibr}}\end{bmatrix}, \Sigma_{\hat{\delta}^{\text{calibr}}}\right\}$$

Matched-pair cluster randomized design



<u>level 1':</u>

$$\begin{bmatrix} \hat{\delta}_1^{\text{calibr}} \\ \vdots \\ \hat{\delta}_N^{\text{calibr}} \end{bmatrix} \Big| \begin{bmatrix} \delta_1^{\text{calibr}} \\ \vdots \\ \delta_N^{\text{calibr}} \end{bmatrix}, \theta, \Sigma_{\hat{\delta}^{\text{calibr}}}$$

$$\sim Normal \left\{ \begin{bmatrix} \delta_1^{\text{calibr}} \\ \vdots \\ \delta_N^{\text{calibr}} \end{bmatrix}, \Sigma_{\hat{\delta}^{\text{calibr}}} \right\}$$

<u>level 2':</u> $\delta_p^{\text{calibr}} \mid \delta^{\text{effect}}, \tau^2 \sim Normal(\delta^{\text{effect}}, \tau^2),$

$$p = 1, \ldots, N.$$

Checking second-level dependence