# Exploring New Statistical Methods for Causal Inference in Longitudinal Studies

Zhiqiang Tan

Department of Statistics
Rutgers University

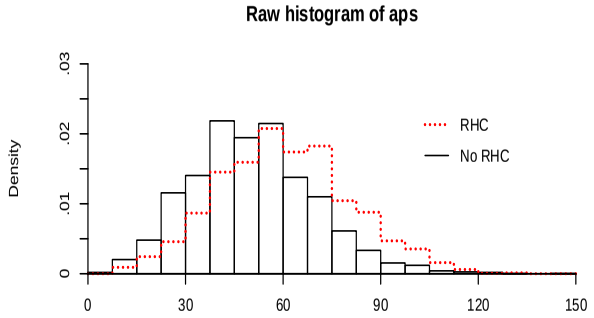
June 8, 2012
Spring Symposium of New Jersey Chapter of ASA

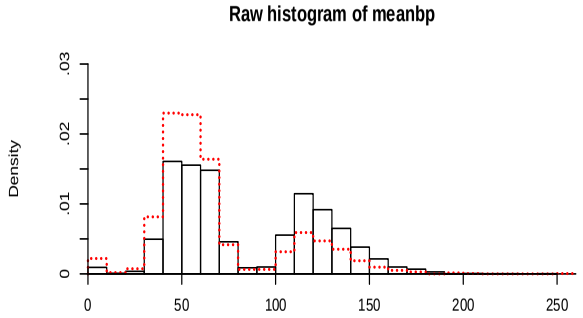
## Outline

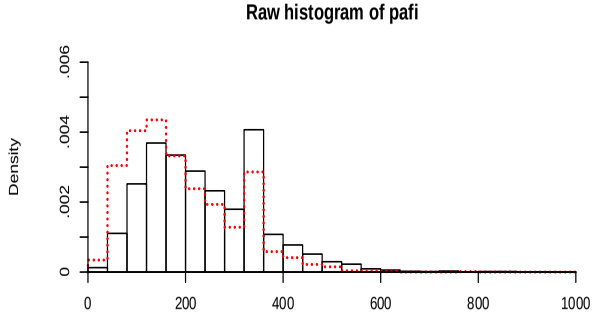
- An introduction: effect of right heart catheterization for critically ill patients
- Review of causal inference in cross-sectional studies
  - Outcome regression
  - Propensity score
- Causal inference in longitudinal studies
  - "Realized" structural models
  - Generalized Mantel-Haenszel analysis
- An application: effect of antipsychotic treatment for elderly nursing home residents

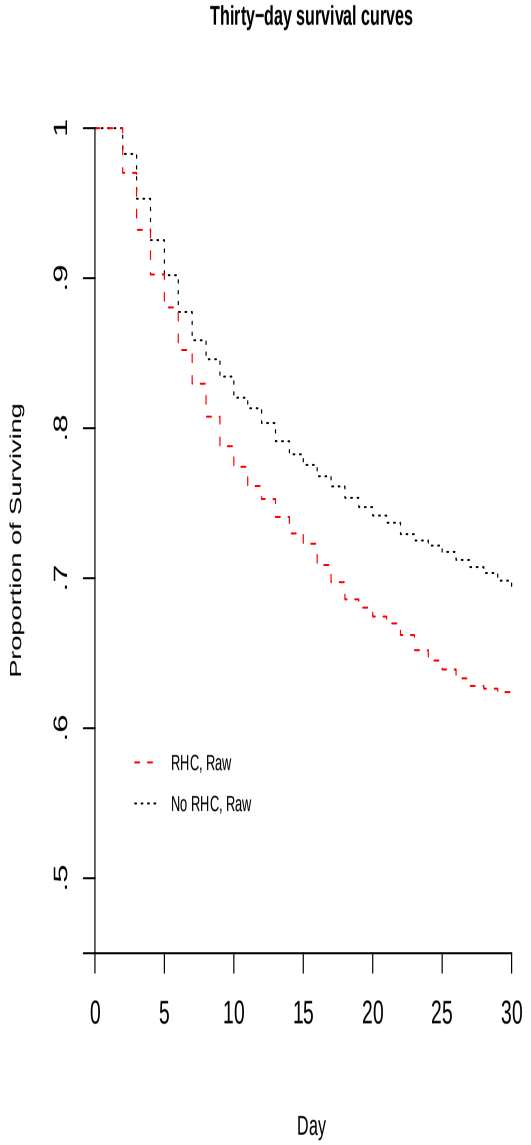
**An introduction: effect of right heart catheterization for critically ill patients**

- Right heart catheterization (RHC) is performed daily in hospitals since 1970s.
- The benefit of RHC had not been demonstrated in a successful randomized clinical trial, until $\geq$ 1990s.
- Connors et al.'s (1996) observational study raised the concern: RHC might not benefit patients and might in fact cause harm.
- Data were collected on 5735 critically ill patients in ICUs:
  - Treatment: No-RHC or RHC
  - Outcome: 30-day survival
  - Covariates: 75 covariates (specified by doctors)
- What is the effect of RHC on survival?


Raw histogram of aps
Density
0
.01
.02
.03
0
30
60
90
120
150
RHC
No RHC


Raw histogram of meanbp
Density
0
.01
.02
.03
0
50
100
150
200
250


Raw histogram of pafi
Density
0
.002
.004
.006
0
200
400
600
800
1000


Thirty-day survival curves
Proportion of Surviving
.5
.6
.7
.8
.9
1
0
5
10
15
20
25
30
Day
RHC, Raw
No RHC, Raw

## Causal inference in cross-sectional studies

- $X$: pre-treatment covariates
- $A$: treatment variable taking value 0 or 1
  if a patient actually received No-RHC or RHC
- $\{Y(0), Y(1)\}$: potential outcome that would be observed
  had a patient received No-RHC or RHC
- $Y$: observed outcome
- Consistency assumption:

$$Y = \begin{cases} Y(0) & \text{if } A = 0 \\ Y(1) & \text{if } A = 1 \end{cases}$$

## Data structure

| id | $X_1$ | $\cdots$ | $X_{75}$ | $A$ | $Y$ | $Y_0$ | $Y_1$ |
|---|---|---|---|---|---|---|---|
| 1 | * | * | * | 0 | 10 | 10 | ?? |
| 2 | * | * | * | 1 | 7 | ?? | 7 |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| N | * | * | * | 0 | 12 | 12 | ?? |

## (Average) Causal effects

- Comparisons of potential outcomes under different treatments
- Causal risk difference

$$\begin{aligned} & E\{Y(1)\} - E\{Y(0)\} \\ =& P\{Y(1) = 1\} - P\{Y(0) = 1\} \\ \neq& P(Y = 1|A = 1) - P(Y = 1|A = 0) \end{aligned}$$

- Causal odds ratio

$$\begin{aligned} & \frac{P\{Y(1) = 1\}\ P\{Y(0) = 0\}}{P\{Y(1) = 0\}\ P\{Y(0) = 1\}} \\ \neq& \frac{P(Y = 1|A = 1)\ P(Y = 0|A = 0)}{P(Y = 0|A = 1)\ P(Y = 1|A = 0)} \end{aligned}$$

**Assumption of no (unmeasured) confounding**

- $\{Y(0), Y(1)\}$ and $A$ are conditionally independent given $X$

$$\{Y(0), Y(1)\} \coprod A \,|\, X$$

- For a randomized clinical trial (RCT),

$$\{Y(0), Y(1)\} \coprod A$$

- Example of "causal algebra":

$$P\{Y(1) = 1\} \overset{independence}{=} P\{Y(1) = 1 | A = 1\}$$
$$\overset{consistency}{=} P(Y = 1 | A = 1)$$

- For an observational study, a challenge is to collect data on a sufficient number of confounding variables such that the assumption holds approximately.

## Statistical approaches under no-confounding

- Consider $\mu_1 = E\{Y(1)\}$ and $\mu_0 = E\{Y(0)\}$. Then

$$\text{causal risk difference} = \mu_1 - \mu_0$$

$$\text{causal odds ratio} = \frac{\mu_1(1-\mu_0)}{(1-\mu_1)\mu_0}$$

- Outcome regression:

$$\text{logit } P(Y=1|A,X) = \alpha_0 + \alpha_1 A + \alpha_2^{\mathrm{T}} X.$$

- Let $(\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2)$ be the standard estimators. Then

$$\hat{\mu}_1^{\mathrm{OR}} = \frac{1}{N}\sum_{i=1}^{N} \text{expit}(\hat{\alpha}_0 + \hat{\alpha}_1 + \hat{\alpha}_2^{\mathrm{T}} X_i),$$

$$\hat{\mu}_0^{\mathrm{OR}} = \frac{1}{N}\sum_{i=1}^{N} \text{expit}(\hat{\alpha}_0 + \hat{\alpha}_2^{\mathrm{T}} X_i).$$

## Statistical approaches under no-confounding (cont'd)

- Propensity score:

$$\text{logit } P(A=1|X) = \gamma_0 + \gamma_1^{\mathrm{T}} X.$$

- Let $(\hat{\gamma}_0, \hat{\gamma}_1)$ be the standard estimators. The fitted propensity score is $\hat{\pi}(X) = \text{expit}(\hat{\gamma}_0 + \hat{\gamma}_1^{\mathrm{T}} X)$.

- Two estimators of $\mu_1$ are

$$\hat{\mu}_1^{\text{IPW}} = \frac{1}{N}\sum_{i=1}^{N} \frac{A_i}{\hat{\pi}(X_i)} Y_i, \quad \hat{\mu}_1^{\text{IPW,ratio}} = \sum_{i=1}^{N} \frac{A_i}{\hat{\pi}(X_i)} Y_i \Big/ \sum_{i=1}^{N} \frac{A_i}{\hat{\pi}(X_i)}.$$

  Two estimators of $\mu_0$ are

$$\hat{\mu}_0^{\text{IPW}} = \frac{1}{N}\sum_{i=1}^{N} \frac{1-A_i}{1-\hat{\pi}(X_i)} Y_i, \quad \hat{\mu}_0^{\text{IPW,ratio}} = \sum_{i=1}^{N} \frac{1-A_i}{1-\hat{\pi}(X_i)} Y_i \Big/ \sum_{i=1}^{N} \frac{1-A_i}{1-\hat{\pi}(X_i)}.$$

- Improved estimators of $(\mu_1, \mu_0)$ are available (Tan 2006, 2010).

## Inverse propensity-score weighting

- Each treated individual $i$ is weighted by

$$w_i^1 = \frac{1}{N} \times \frac{1}{\hat{\pi}(X_i)}.$$

  The estimator of $\mu_1$ is $\sum_{i:A_i=1} w_i^1 \, Y_i$.

- Each untreated individual $i$ is weighted by

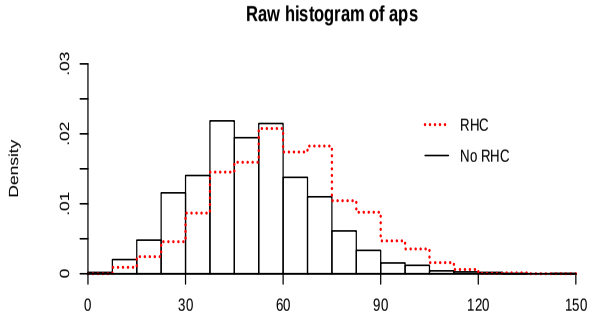
$$w_i^0 = \frac{1}{N} \times \frac{1}{1 - \hat{\pi}(X_i)}.$$

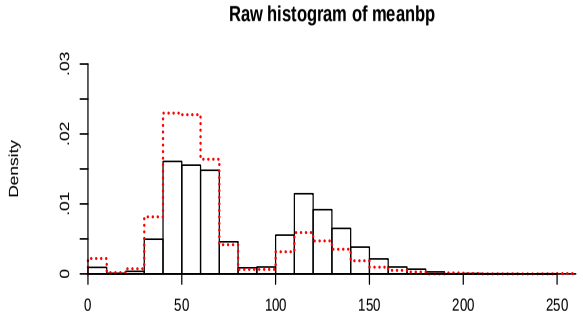
  The estimator of $\mu_0$ is $\sum_{i:A_i=0} w_i^0 \, Y_i$.

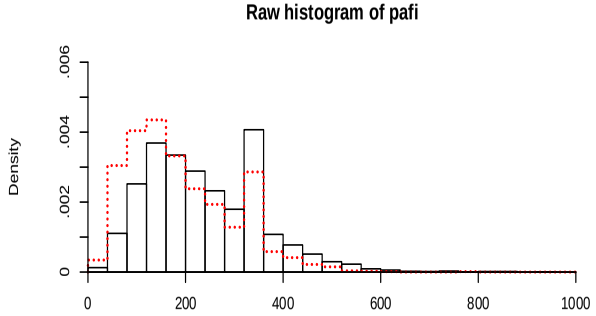
- Covariate balance:

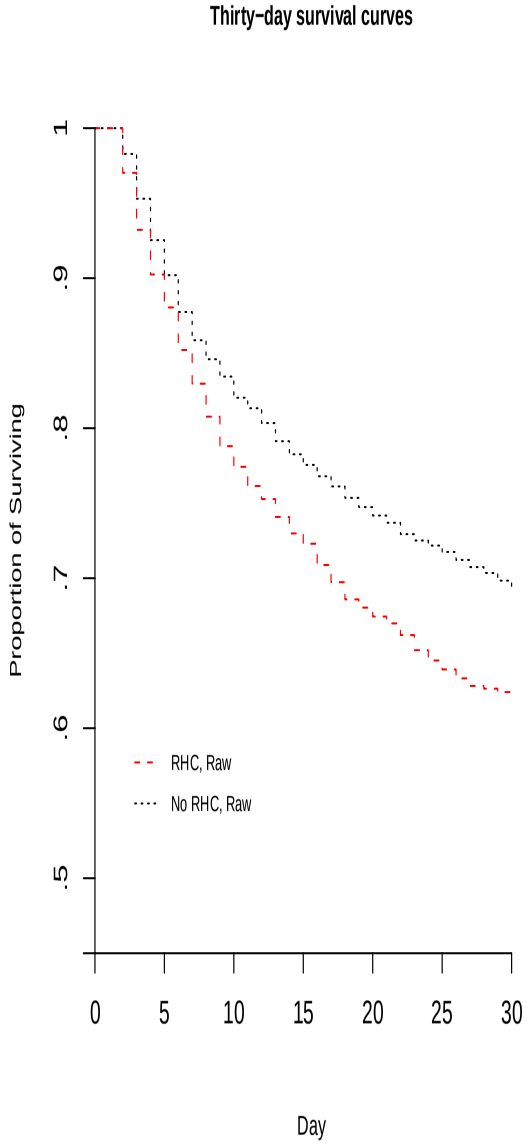
$$\sum_{i:A_i=1} w_i^1 \, \varphi(X_i) \approx \sum_{i:A_i=0} w_i^0 \, \varphi(X_i)$$

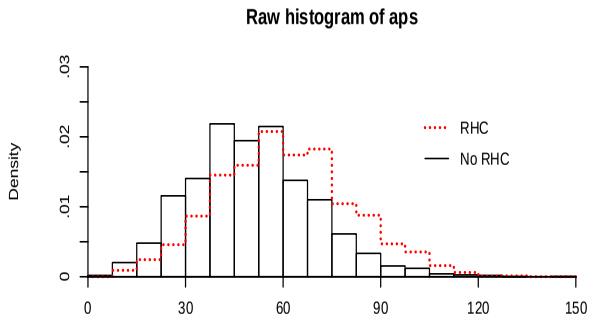
  for an arbitrary function $\varphi(\cdot)$.

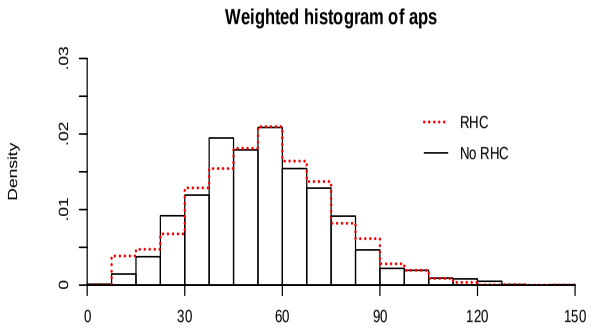

Raw histogram of aps
Density
0
.01
.02
.03
0
30
60
90
120
150
RHC
No RHC

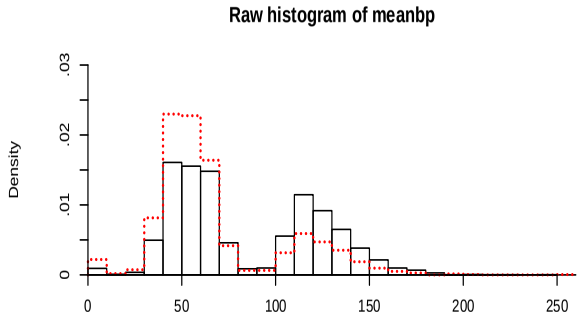

Raw histogram of meanbp
Density
0
.01
.02
.03
0
50
100
150
200
250

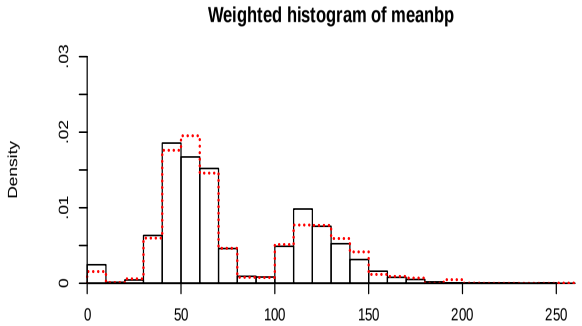

Raw histogram of pafi
Density
0
.002
.004
.006
0
200
400
600
800
1000

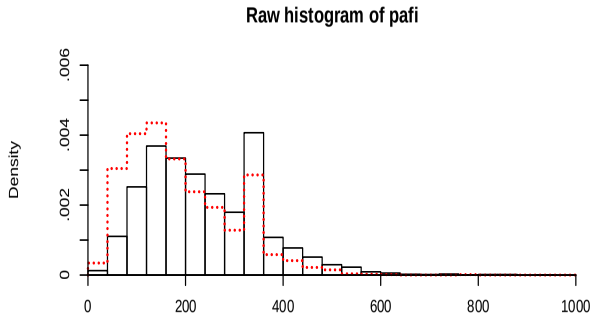

Thirty-day survival curves
Proportion of Surviving
.5
.6
.7
.8
.9
1
0
5
10
15
20
25
30
Day
RHC, Raw
No RHC, Raw

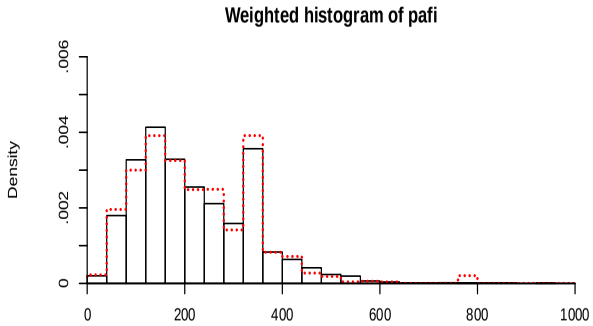

Raw histogram of aps
Density
RHC
No RHC
0
.01
.02
.03
0
30
60
90
120
150

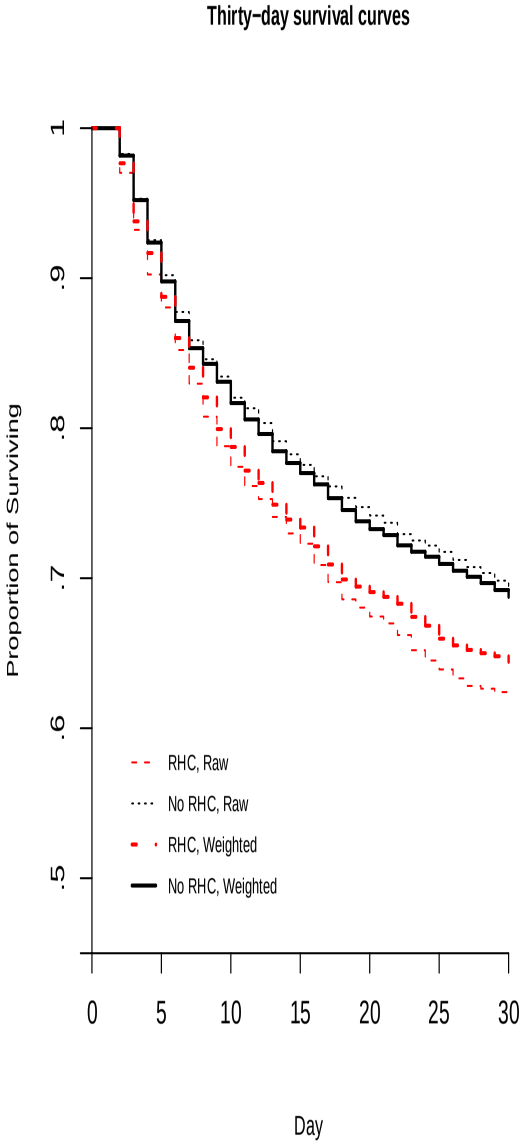

Weighted histogram of aps
Density
RHC
No RHC
0
.01
.02
.03
0
30
60
90
120
150


Raw histogram of meanbp
Density
0
.01
.02
.03
0
50
100
150
200
250


Weighted histogram of meanbp
Density
0
.01
.02
.03
0
50
100
150
200
250


Raw histogram of pafi
Density
0
.002
.004
.006
0
200
400
600
800
1000


Weighted histogram of pafi
Density
0
.002
.004
.006
0
200
400
600
800
1000


Thirty-day survival curves
Proportion of Surviving
.5
.6
.7
.8
.9
1
RHC, Raw
No RHC, Raw
RHC, Weighted
No RHC, Weighted
0
5
10
15
20
25
30
Day

## Causal inference in longitudinal studies

- The temporally ordered observed data are

$$(L_0,\ A_1,\ L_1,\ A_2,\ L_2,\ ...,\ A_{K-1},\ L_{K-1},\ A_K,\ L_K).$$

  - $L_0$: baseline variables
  - $A_k$: treatment variable (0/1) for the $k$th time interval
  - $L_k = (X_k, Y_k)$ where

    $X_k$: confounding variables for the $k$th time interval

    $Y_k$: outcome variable for the $k$th time interval

    [ 0 if alive at the end of the $k$th interval or 1 otherwise ]

- The history up to and including $k$th interval

$$\bar{A}_k = (A_1, \ldots, A_k),$$
$$\bar{L}_k = (L_0, L_1, \ldots L_k).$$

## Causal inference in longitudinal studies (cont'd)

- For $\bar{a}_k = (a_1, \ldots, a_k)$, define

$$Y_k(\bar{a}_k) = Y_k(a_1, \ldots, a_k)$$

  as the potential outcome that would be observed at the end of $k$th interval if the individual followed the treatment $\bar{a}_k$.

- Consistency assumption:

$$Y_k = Y_k(\bar{a}_k) \quad \text{if} \quad \bar{A}_k = \bar{a}_k.$$

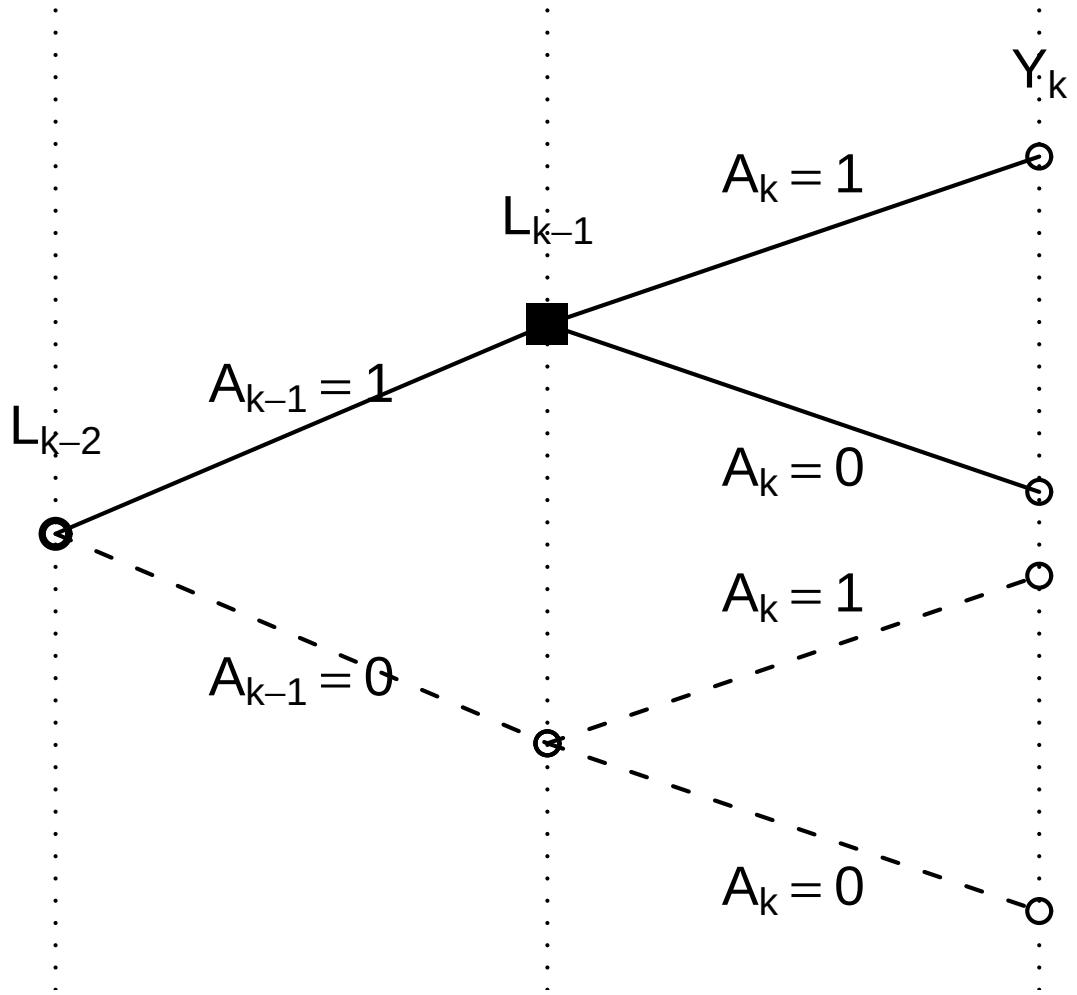
- Assumption of no (unmeasured) confounding:

$$Y_k(\bar{a}_k) \coprod A_k \,|\, \bar{A}_{k-1} = \bar{a}_{k-1}, \bar{L}_{k-1}.$$

  For $k = 2$,

$$Y_2(a_1, a_2) \coprod A_2 \,|\, A_1 = a_1, L_0, L_1.$$

## Longitudinal data structure




Potential data grow exponentially!

## Marginal structural models

- A marginal structural logistic model (Hernan et al. 2000)

$$\text{logit } P\{Y_k(\bar{a}_k) = 1 | Y_{k-1}(\bar{a}_{k-1}) = 0\} = \theta_0(k) + \theta a_k,$$

  where $\theta_0(k)$ is a time-specific intercept and
  $\theta$ is causal log odds ratio, constant for all $k$ and $\bar{a}_{k-1}$.

- The conditioning event $\{Y_{k-1}(\bar{a}_{k-1}) = 0\}$ represents

  *the subset of all individuals in the source population who would survive at least $k-1$ periods if they followed the treatment $\bar{a}_{k-1}$.*

- This subset cannot be directly identified from observed data.

## Estimation for marginal structural models

- Weights based on sequential propensity scores:

$$SW(k) = \prod_{j=1}^{k} \frac{\pi(A_j|\bar{A}_{j-1})}{\pi(A_j|\bar{A}_{j-1}, \bar{L}_{j-1})},$$

  where $\pi(a_j|\bar{a}_{j-1}, \bar{l}_{j-1}) = P(A_j = a_j|\bar{A}_{j-1} = \bar{a}_{j-1}, \bar{L}_{j-1} = \bar{l}_{j-1})$.

- Assumption: any individual starting treatment remained on treatment thereafter, i.e., if $A_{k-1} = 1$ then $A_k = 1$.

- Implementation
  - Fit a pooled logistic regression model

$$\text{logit } P(A_k = 0|A_{k-1} = 0, \bar{L}_{k-1}) = \gamma_0(k) + \gamma_1^{\mathrm{T}} L_{k-1}.$$

    Let $\widehat{SW}(k)$ be the resulting estimator of $SW(k)$.
  - Fit the model

$$\text{logit } P(Y_k = 1|Y_{k-1} = 0, \bar{A}_k) = \theta_0(k) + \theta A_k$$

    by weighted pooled logistic regression using weights $\widehat{SW}(k)$.

## "Realized" structural models

- Consider the structural model

  $$\text{logit } P\{Y_k(\bar{a}_k) = 1 | Y_{k-1}(\bar{a}_{k-1}) = 0, \bar{A}_{k-1} = \bar{a}_{k-1}\} = \theta_0(\bar{a}_{k-1}) + \theta_k a_k,$$

  where the intercept $\theta_0(\bar{a}_{k-1})$ is allowed to depend on $\bar{a}_{k-1}$ and the slope $\theta_k$ is allowed to depend on $k$, but not on $\bar{a}_{k-1}$.

- The conditioning event $\{Y_{k-1}(\bar{a}_{k-1}) = 0, \bar{A}_{k-1} = \bar{a}_{k-1}\}$ equals

  $$C(\bar{a}_{k-1}) = \{Y_{k-1} = 0, \bar{A}_{k-1} = \bar{a}_{k-1}\},$$

  representing

  *the subset of individuals who actually followed treatment $\bar{a}_{k-1}$ and survived at least $k-1$ periods.*

- This subset can be directly identified from observed data.

## "Realized" structural models (cont'd)

- The parameter $\theta_k$ gives

$$e^{\theta_k} = \frac{P\{Y_k(\bar{a}_{k-1}, 1) = 1 | C(\bar{a}_{k-1})\}\ P\{Y_k(\bar{a}_{k-1}, 0) = 0 | C(\bar{a}_{k-1})\}}{P\{Y_k(\bar{a}_{k-1}, 1) = 0 | C(\bar{a}_{k-1})\}\ P\{Y_k(\bar{a}_{k-1}, 0) = 1 | C(\bar{a}_{k-1})\}}.$$

- This is exactly the cross-sectional causal odds ratio for the "population" $C(\bar{a}_{k-1})$, with pre-treatment covariates $\bar{L}_{k-1}$, treatment $A_k$, and outcome $Y_k$.
- Consistent estimators of $\theta_k$ can be obtained ...
  - by applying the cross-sectional methods to the subsample of individuals restricted to $C(\bar{a}_{k-1})$,
  - provided the subsample size is sufficiently large, say $\geq 500$.

**Outcome regression and propensity scores** for each $a_{k-1}$

- An outcome regression model is

$$\text{logit } P\{Y_k = 1 | A_k, \bar{L}_{k-1}, C(\bar{a}_{k-1})\} = \alpha_0(\bar{a}_{k-1}) + \alpha_1(\bar{a}_{k-1})A_k + \alpha_2^{\mathrm{T}}(\bar{a}_{k-1})L_{k-1},$$

  where the parameters $\alpha_0(\bar{a}_{k-1})$, $\alpha_1(\bar{a}_{k-1})$, and $\alpha_2(\bar{a}_{k-1})$ are allowed to depend on $\bar{a}_{k-1}$.

- A propensity score model is

$$\text{logit } P\{A_k = 1 | \bar{L}_{k-1}, C(\bar{a}_{k-1})\} = \gamma_0(\bar{a}_{k-1}) + \gamma_1^{\mathrm{T}}(\bar{a}_{k-1})L_{k-1},$$

  where the parameters $\gamma_0(\bar{a}_{k-1})$ and $\gamma_1(\bar{a}_{k-1})$ are allowed to depend on $\bar{a}_{k-1}$.

## Combining data across treatment trajectories

- So far, our discussion is to obtain consistent estimators of $\theta_k$ using the subsample restricted to $C(\bar{a}_{k-1})$ for each treatment trajectory $\bar{a}_{k-1}$ subject to the subsample size requirement.
- The assumption that the causal odds ratio is constant across different $\bar{a}_{k-1}$ allows us to combine such estimators of $\theta_k$ based on subsamples corresponding to different $\bar{a}_{k-1}$.
- This is similar to the Mantel-Haenszel (1959) estimator of the common odds ratio in a series of $2 \times 2$ tables.
- If the variables $L_{k-1}$ are dropped, then the Mantel-Haenszel estimator can be directly applied to estimate $\theta_k$, using a $2 \times 2$ table for the subsample restricted to $C(\bar{a}_{k-1})$ for each $\bar{a}_{k-1}$.

  $\hookrightarrow$ the crude estimator of $\theta_k$, without adjusted for any time-dependent variables.

## $2 \times 2$ table for Mantel-Haenszel estimation

| | $Y_k = 0$ | $Y_k = 1$ |
|---|---|---|
| $A_k = 0$ | $n_{00}$ | $n_{01}$ |
| $A_k = 1$ | $n_{10}$ | $n_{11}$ |

$n_{ay}$ is the number of individuals with $(A_k, Y_k) = (a, y)$ in the subsample restricted to $C(\bar{a}_{k-1})$.

## Generalized Mantel-Haenszel methods with time-dependent variables

- For each $\bar{a}_{k-1}$, let

  $\hat{\mu}_1(\bar{a}_{k-1})$ be an estimator of $P\{Y_k(\bar{a}_{k-1}, 1) = 1 | C(\bar{a}_{k-1})\}$ and

  $\hat{\mu}_0(\bar{a}_{k-1})$ be an estimator of $P\{Y_k(\bar{a}_{k-1}, 0) = 1 | C(\bar{a}_{k-1})\}$.

- The generalized Mantel-Haenszel estimator of $\mathrm{e}^{\theta_k}$ is

$$\frac{\sum_{\bar{a}_{k-1}} \varrho(\bar{a}_{k-1}) \hat{\mu}_1(\bar{a}_{k-1}) \{1 - \hat{\mu}_0(\bar{a}_{k-1})\}}{\sum_{\bar{a}_{k-1}} \varrho(\bar{a}_{k-1}) \{1 - \hat{\mu}_1(\bar{a}_{k-1})\} \hat{\mu}_0(\bar{a}_{k-1})},$$

  where $\sum_{\bar{a}_{k-1}}$ is taken over $a_{k-1}$ subject to the subsample size requirement.

- The factor $\varrho(\bar{a}_{k-1})$ can be chosen such that if no time-dependent variables are involved, then the original Mantel-Haenszel estimator is recovered.

## Advantages of the proposed methods

- Simplicity, transparency, and flexibility
  - "Realized" structural models are conceptually simple.
  - Estimation involves straightforward extensions of cross-sectional methods.
  - Generalized Mantel-Haenszel analysis is flexible to combine data across treatment trajectories
- The method of estimation for marginal structural models requires constructing individual weights by multiplying propensity scores from successive periods.

**An application: effect of antipsychotic treatment for elderly nursing home residents**

- Study cohort: nursing home residents, 65 years or older
  - 114,369 subjects at baseline
  - 45,691, 20,195, and 7,038 subjects remaining in the cohort after 4, 8, and 12 quarters
- Time interval: 3 months (quarter)
- Treatment: 1=antipsychotic medication, 0=otherwise
- Outcome: 1=death from all causes, 0=otherwise
- 51 time-dependent variables: assessments of physical and psychological functioning and active clinical diagnoses

**Analyses and results**

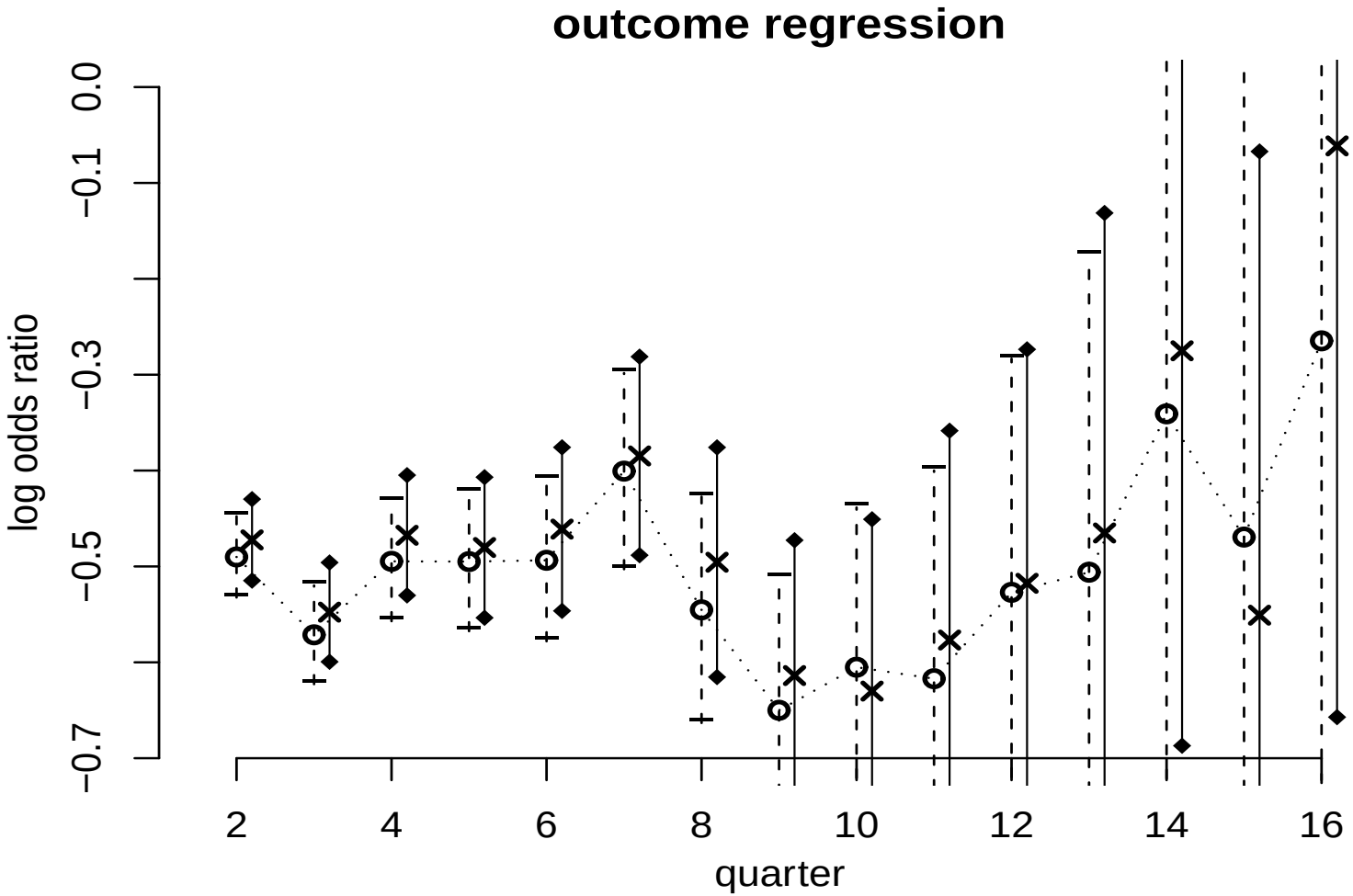
- Crude (unadjusted) Mantel-Haenszel analysis
  - The number of treatment trajectories up to the $k$th quarter, subject to the size requirement,
    ..... initially increases (to 8 for $k = 10$)
    ..... but then quickly decreases (to 3, 1 for $k = 12, 14$)
    as $k$ increases.
- Generalized Mantel-Haenszel analysis, taking account of the time-dependent variables
  - Outcome regression method
  - Propensity score method

## $2 \times 2$ tables for the first 4 quarters

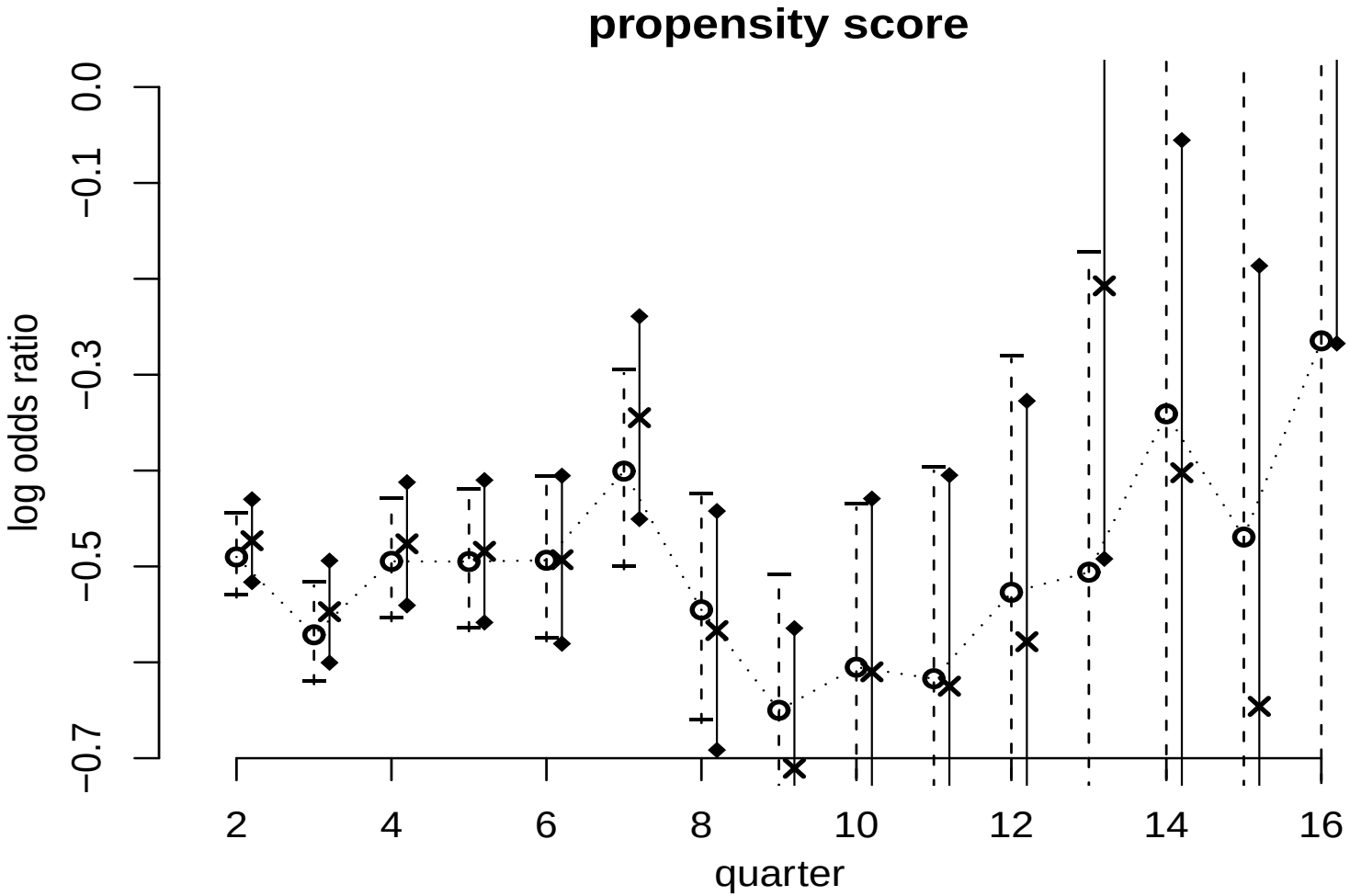
| $k = 1$ | | $k = 2$ | | | $k = 3$ | | | $k = 4$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| tx | death | tx | death | alive | tx | death | alive | tx | death | alive |
| 1 | 32745 | 10 | 3966 | 12127 | 100 | 1919 | 8283 | 1000 | 1376 | 6216 |
| | | 11 | 10512 | 52472 | 101 | 338 | 1587 | 1001 | 124 | 567 |
| | | | | | 110 | 2229 | 5983 | 1010 | 82 | 249 |
| | | | | | 111 | 7049 | 37211 | 1011 | 226 | 1030 |
| | | | | | | | | 1100 | 922 | 4019 |
| | | | | | | | | 1101 | 179 | 863 |
| | | | | | | | | 1110 | 1262 | 3512 |
| | | | | | | | | 1111 | 5154 | 27283 |

Considered are only the tables with the total number of individuals $\geq 500$.

Estimated log odds ratios of death (∘: unadjusted, ×: adjusted) and 95% confidence intervals



Estimated log odds ratios of death (∘: unadjusted, ×: adjusted) and 95% confidence intervals



## Discussion

- The unadjusted analysis showed a reduction in mortality for patients treated with an anti-psychotic medication.

  Adjustment for available time-dependent confounders did not meaningfully alter this estimate.

- On the other hand, meta analysis of 17 short-term randomized controlled trials indicates a 60-70% increased risk of death versus placebo in the elderly population (FDA 2010).

- A possible explanation is that some important time-dependent confounders might not be included in the study.

- The soundness of causal inference depends on the quality of both the data and the methods.