Part II – Structural Mixed Models

- Patrick J. Heagerty
- Department of Biostatistics
- University of Washington

Session Two Outline

- Examples
- Structural Model
- Selection Model for Timing of Surgery
- Analysis Options
- Evaluation of Methods
- Concluding Comments
- Extension to Treatment Selection

Examples

- SPORT Weinstein et al. (2006) $JAMA$
  - Spine Patient Outcomes Research Trial
  - Disk herniation
  - N=501 subjects
  - SF-36 Physical Function assessment through 24 months.
  - **Surgery**:
    - At 3mo: **50%** / **30%**
    - At 12mo: **60%** / **45%**

Examples

- INVEST Kallmes et al. (2009) *NEJM*
  - Vertebroplasty
  - N=131 subjects
  - Assessment at 0, 1, 3, 6, 12 mo
  - **Crossover**:
    - From Tx=A to Tx=B: **11/68**
    - From Tx=B to Tx=A: **32/63**

Is a Common Medical Procedure Unnecessary?

Two Studies Show that No Difference For Patients Who Had Vertebroplasty and Those Who Had Placebo Treatment

By Jonathan LaPook

A A Font size | Print | Share | 39 Comments

VIDEO

Unnecessary Procedures

Studies have shown with those who undergo a common back surgery, vertebroplasty, felt no different. As Dr. Jon LaPook reports, many question if expensive treatments like these really work.

(CBS) Sister Rogene Fox, 81, was suffering from severe back pain until she agreed to a popular treatment, reports **CBS News medical correspondent Dr. Jon LaPook**. She believed it worked.

"I just thought, thank God. I don't care what I received," Fox said. "I feel good! I don't have pain!"

But it turns out she got relief without getting the procedure, called Vertebroplasty, a common treatment for patients with painful back fractures from osteoporosis.

Wednesday, two separate studies in the New England Journal of Medicine report there was no difference up to six months later for patients who actually had the procedure and those who had a fake or placebo treatment instead.

"I thought, 'Wow, we're onto something,'" said Dr. David Kallmes, a study author. "We saw many placebo patients get full pain relief. I have no idea why."

During Vertebroplasty, doctors inject medical cement into fractured bone in the back to strengthen the area and reduce pain. The placebo was just a shot to temporarily numb the area. Vertebroplasty is endorsed by multiple medical societies, but the surprising findings may force doctors to rethink the

Original Goal

- As-treated analysis of surgical data requires use of methods appropriate for **endogenous** exposure.
- Patient status post-randomization is a predictor of seeking treatment change (non-adherence).
- **Implication**:

$$E[Y_i(t) \mid Z_i, \text{Surg}_i = t+1] \neq E[Y_i(t) \mid Z_i, \text{Surg}_i > t+1]$$

- **Hypothesis**:
  - ▷ Standard longitudinal data analysis methods will be biased with endogenous exposure.

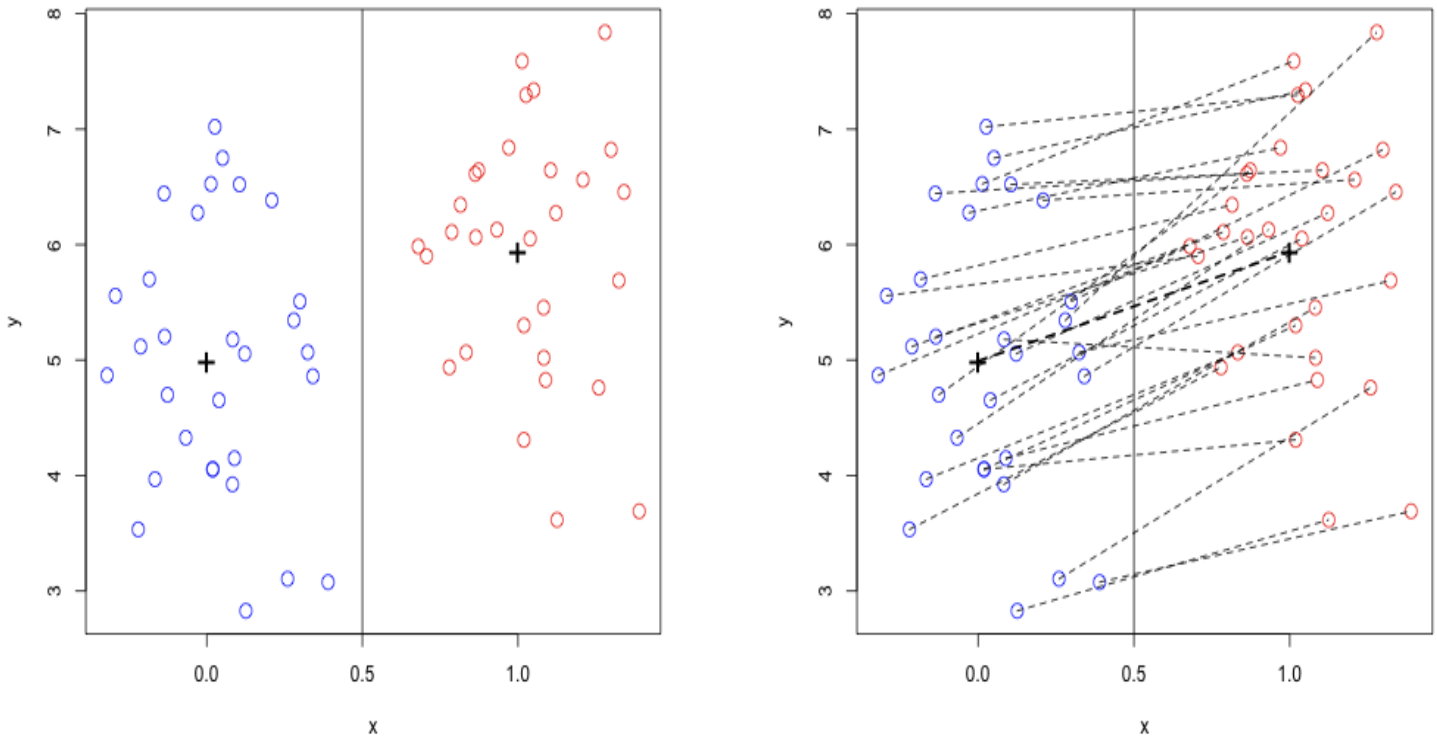
Longitudinal Structural Model

- **Q**: How to determine the "unbiased" estimate when time-of-surgery is a random variable?
- **A**: Use a Longitudinal Structural Mean Model

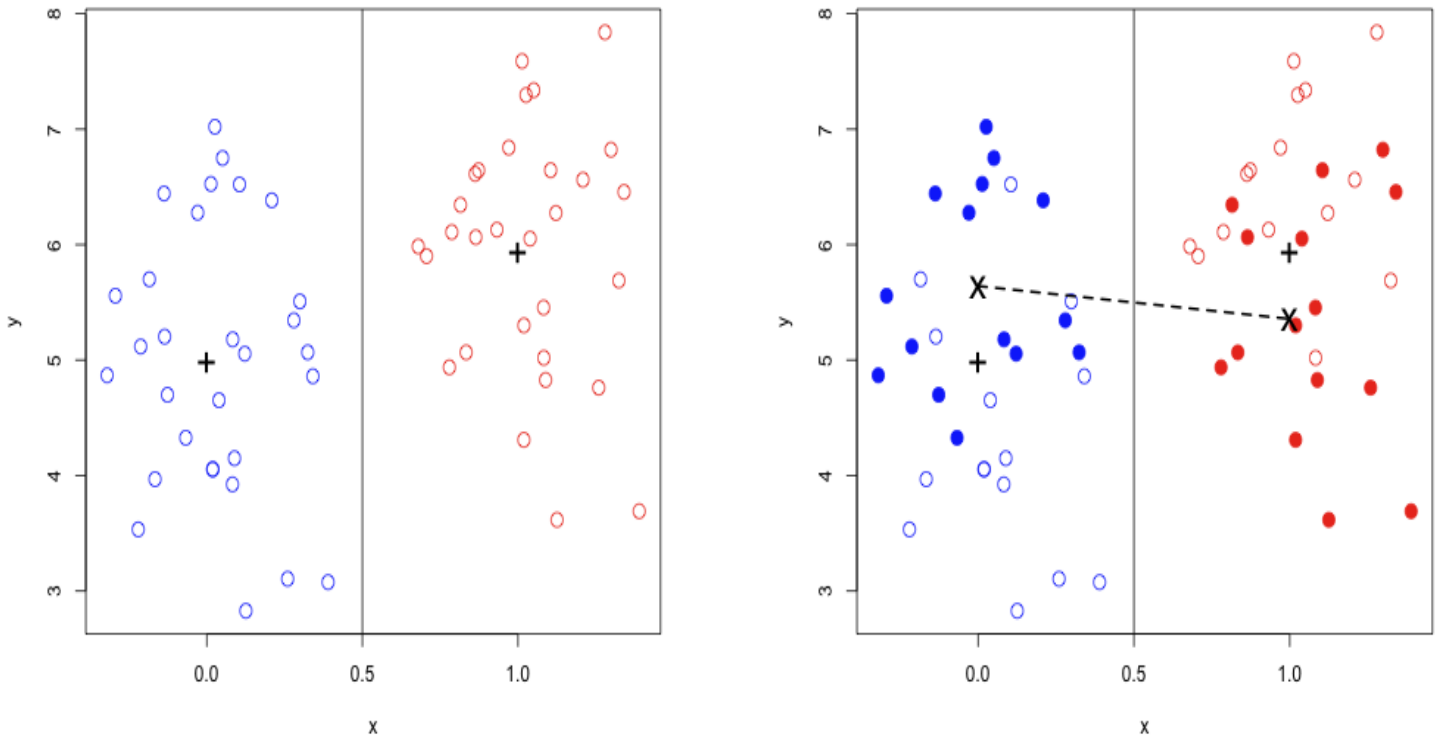
$$Y_i(t,s) \quad = \quad \text{outcome at time } t \text{ when surgery is at time } s$$

$$E[Y_i(t,s)] \quad = \quad \text{population mean when surgery is controlled to be at time } s$$

Basic Structural Model



Basic Structural Model



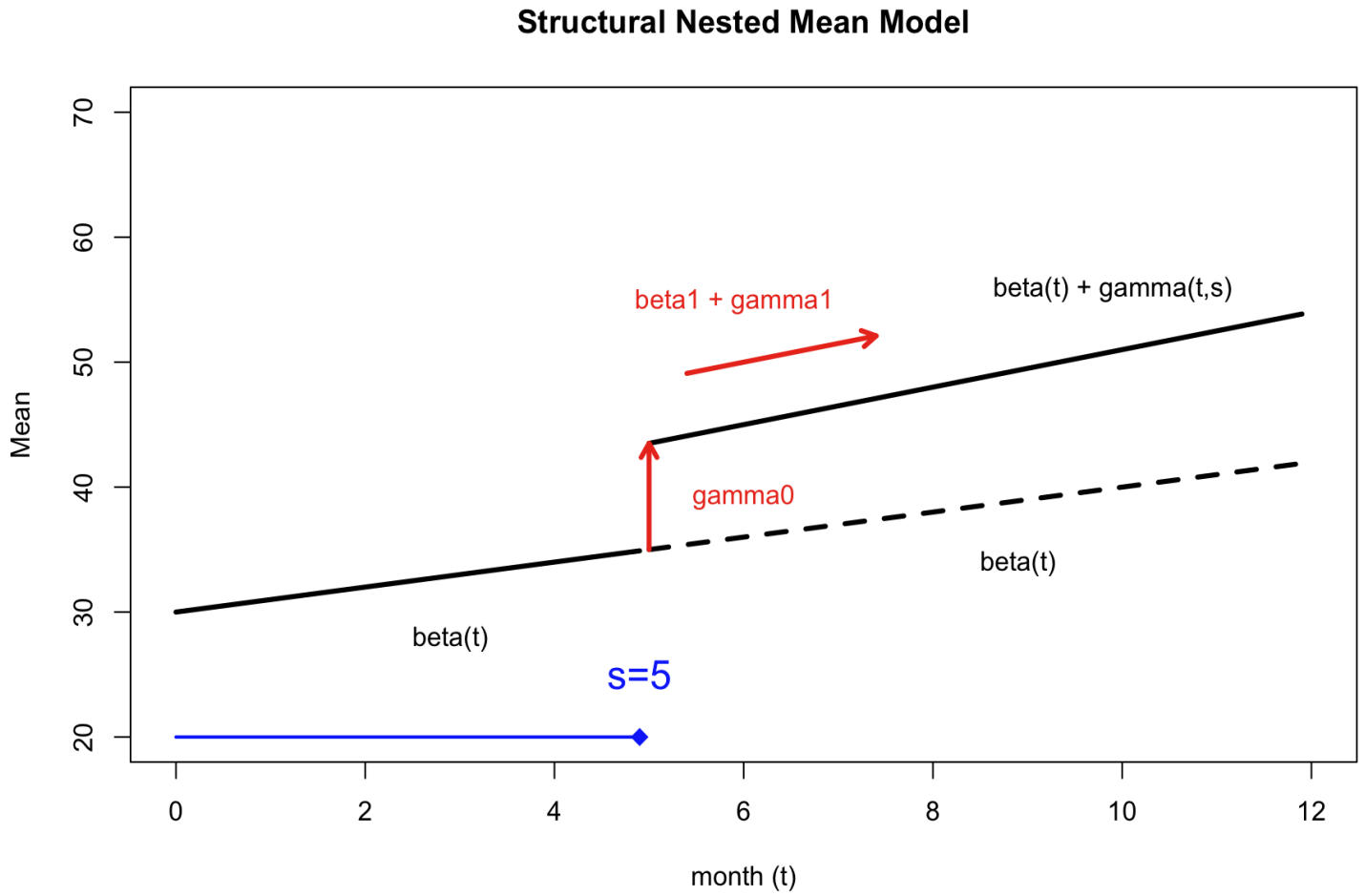
Structural (Nested) Mean Model

- Surgery can be coded as a "point" treatment

	t=1	t=2	t=3	t=4	t=5
s	0	1	0	0	0

- This implies that use of structural nested mean models (Robins 1994) takes a simple form:

$$E[Y_i(t,s)] = \beta(t) + \gamma(t,s) \cdot 1(t > s)$$


Structural Nested Mean Model
Mean
70
60
50
40
30
20
0
2
4
6
8
10
12
month (t)
beta1 + gamma1
beta(t) + gamma(t,s)
gamma0
beta(t)
beta(t)
s=5

Longitudinal Structural Mixed Model

- Data: Z_i=Tx assigned; S_i = surgical time;
  Outcomes = $Y_i(t, S_i)$
- **Q**: How to simulate surgical outcome data with a given causal structure and endogenous surgical timing?
- **A**: Structural Distribution Model

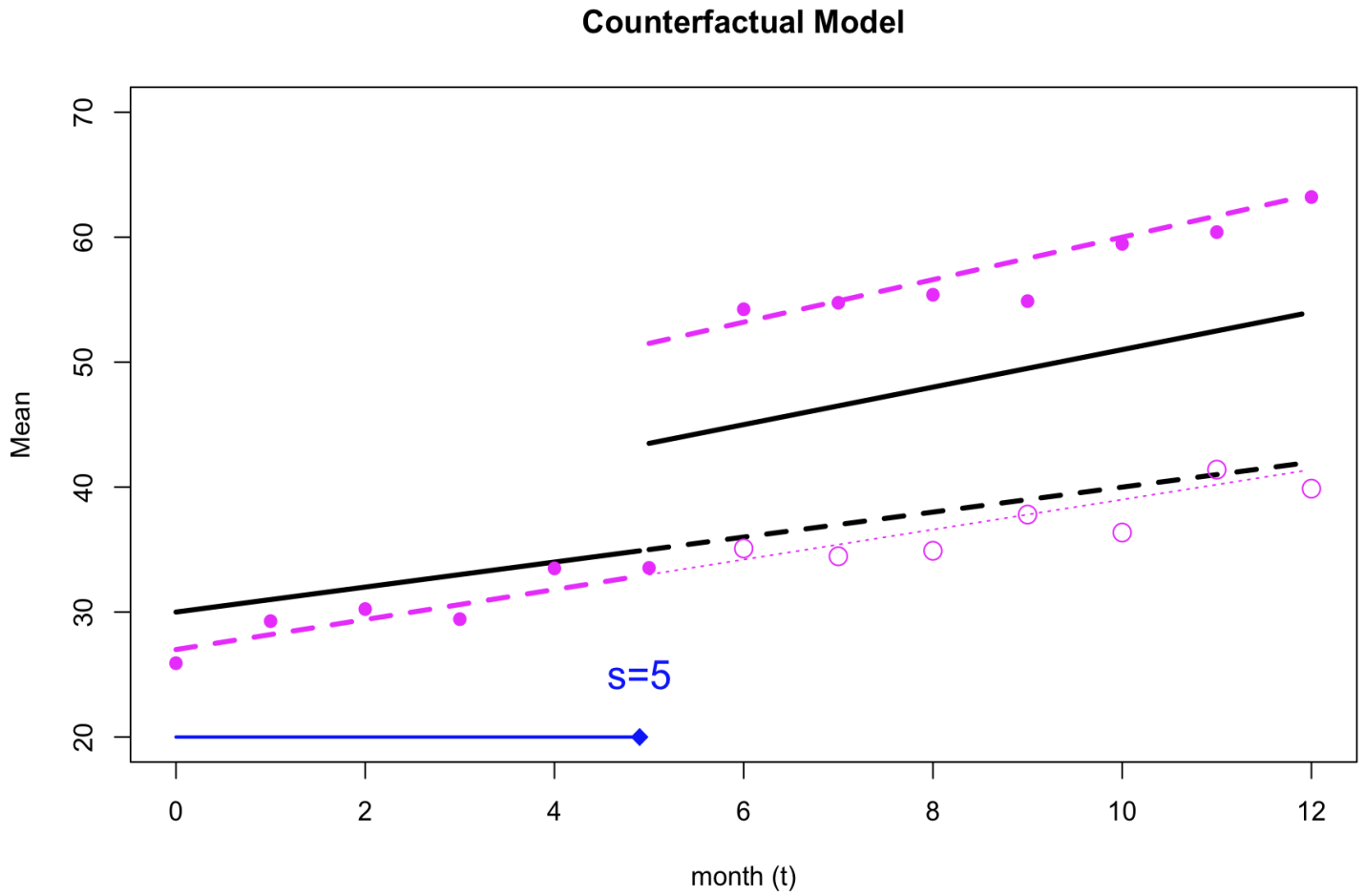
$$
\begin{aligned}
Y_i(t,s) \quad &= \quad \beta(t) \;+\; \gamma(t,s)\cdot 1(t>s) && \text{population} \\
&\quad + \; b_i(s,t) && \text{subject} \\
&\quad + \; e_i(s,t) && \text{observation}
\end{aligned}
$$

Longitudinal Structural Mixed Model

- Simple Example:

$$
\begin{aligned}
Y_i(t,s) \;=\; & \beta_0 + \beta_1 \cdot t \;+\; [\gamma_0 + \gamma_1 \cdot (t-s)] \cdot 1(t > s) \\
& + \; b_{i,0} \;+\; b_{i,1} \cdot t \;+\; b_{i,2} \cdot 1(t > s) \\
& + \; e_{i,0}(t) \cdot 1(t \le s) + \; e_{i,1}(t) \cdot 1(t > s)
\end{aligned}
$$

distribution $\quad b_i \sim \mathcal{N}, \;\; e_i \sim \mathcal{N}$


Counterfactual Model
Mean
70
60
50
40
30
20
0
2
4
6
8
10
12
month (t)
s=5

Selection Model(s) for Surgery

- Define: $Y_i^O(t)$ observed outcome at time t.

$$p_i(s) = P[S_i = s \mid S_i \geq s, Z_i, \{Y_i^O(t)\ t < s\},\ b_i\ ]$$

- **Direct**: (like MAR)

$$p_i(s) = f[\ \ Z_i,\ \ \{Y_i^O(t)\ t < s\}\ \ ]$$

- **Indirect**: (like NMAR)

$$p_i(s) = f[\ \ Z_i,\ \ \{Y_i^O(t)\ t < s\},\ \ b_i]$$

Estimation Options

- **Linear Mixed Models**
- **Generalized Estimating Equations (GEE)**
- **Marginal Structural Models (MSM)**
  - ▷ Selection model; IPW; (R, H& B 2000)
- **g-Estimation**
  - ▷ SNMM; semi-par efficient (Robins 1994)
  - ▷ Extension of IV (Joffe 2004; Dunn 2007)
- **Instrumental Variable (IV) Analysis**
  - ▷ 2SLS; (Woolridge 2002)

Multivariate Structure

- In order to specify a multivariate regression model it is typical to consider:

$$\text{vec}[Y_i^O(t)] \mid \text{vec}[X_i(t)]$$

  where $X_i(t) = [Z_i, 1(t > S_i)]$.

- Endogeneity implies

$$Y_i^O(t) \mid X_i(1), \ldots, X_i(t), \ldots, X_i(T)$$

  depends on $X_i(t+1) \ldots X_i(T)$.

Multivariate Conditional Mean

- Then we have

$$E\left\{Y_i^O(t) \mid \text{vec}[X_i(t)]\right\} \neq E[Y_i^O(t) \mid X_i(t)]$$

$$\neq \beta(t) + \gamma(t,s) \cdot 1(t > s)$$

- Therefore moment-based arguments for regression validity do not apply.
  - ▷ LMM maximum likelihood as WLS
  - ▷ GEE

Marginal Structural Models

- Robbins, Hernan, Brumback (2000)
- General Approach Structural model and...
  - **Selection model** for treatment (e.g. $p_i(s)$ described earlier) used to reweight data in order to correct for selection bias.
  - Semi-parametric since only moments of structural model are specified.
- Key Assumptions
  - Model for treatment is correct.

Instrumental Variable Methods

- Wooldridge (2002)
- General Approach Structural model and **IV**:

$$\begin{aligned} Y &= X\beta + e_i \\ X &= Z\alpha + \epsilon_i \end{aligned}$$

  - $\widehat{\beta}$ solution to: $0 = Z^T(Y - X\beta)$
- Key Assumptions
  - Instrument Z is not correlated with e_i.
  - Randomization indicator, Z, is an IV!

SNMM / G-Estimation

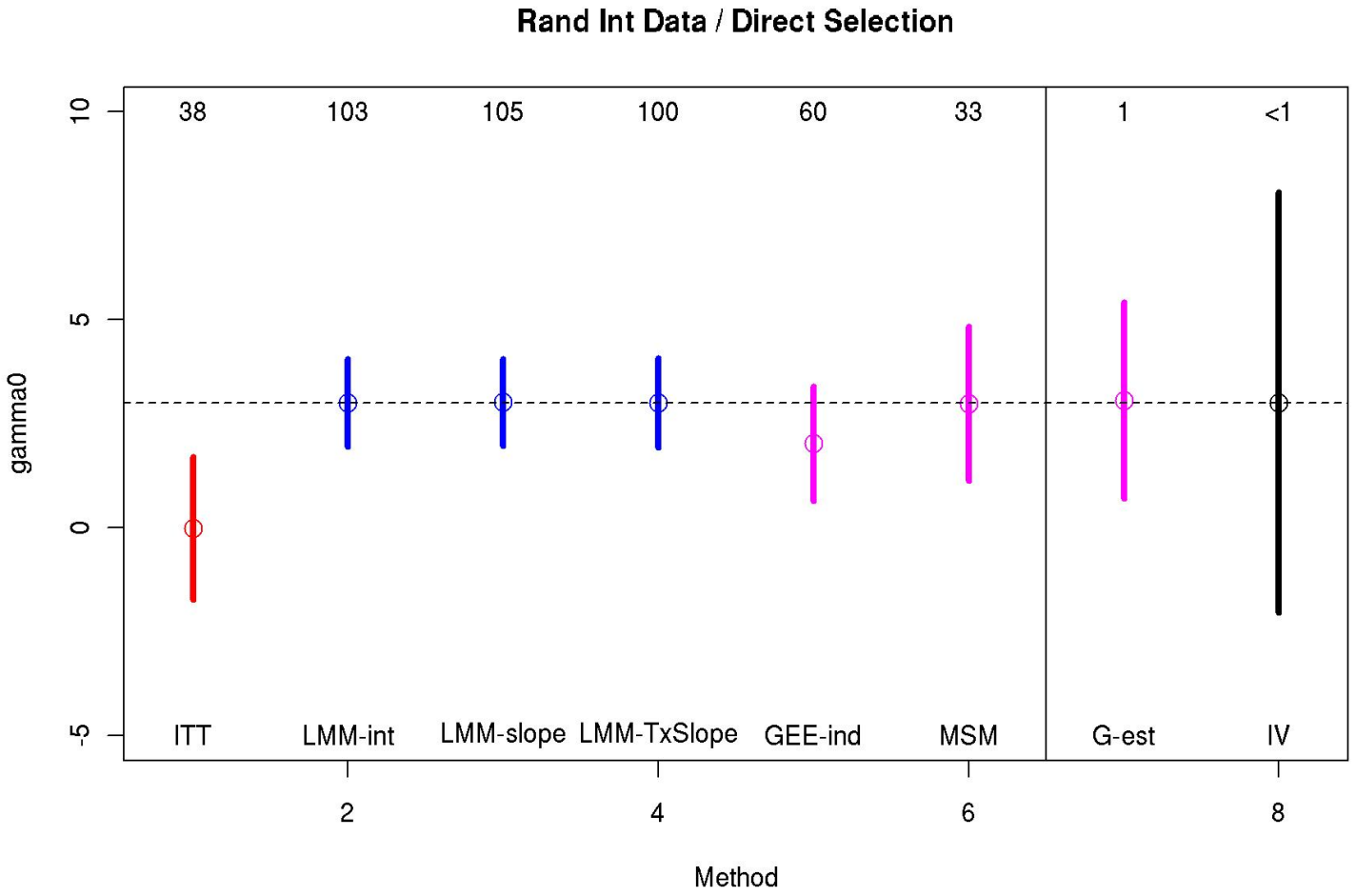
- Robbins (1994)
- General Approach Structural model and...
  - **Estimating equation**
  - Solve: $0 = g(Z)^T W(Z)[Y - \mu(X, Z)]$
  - Semi-parametric since only moments of structural model are specified.
  - Special case is IV estimator.
- Key Assumptions
  - Typically "sequential randomization".

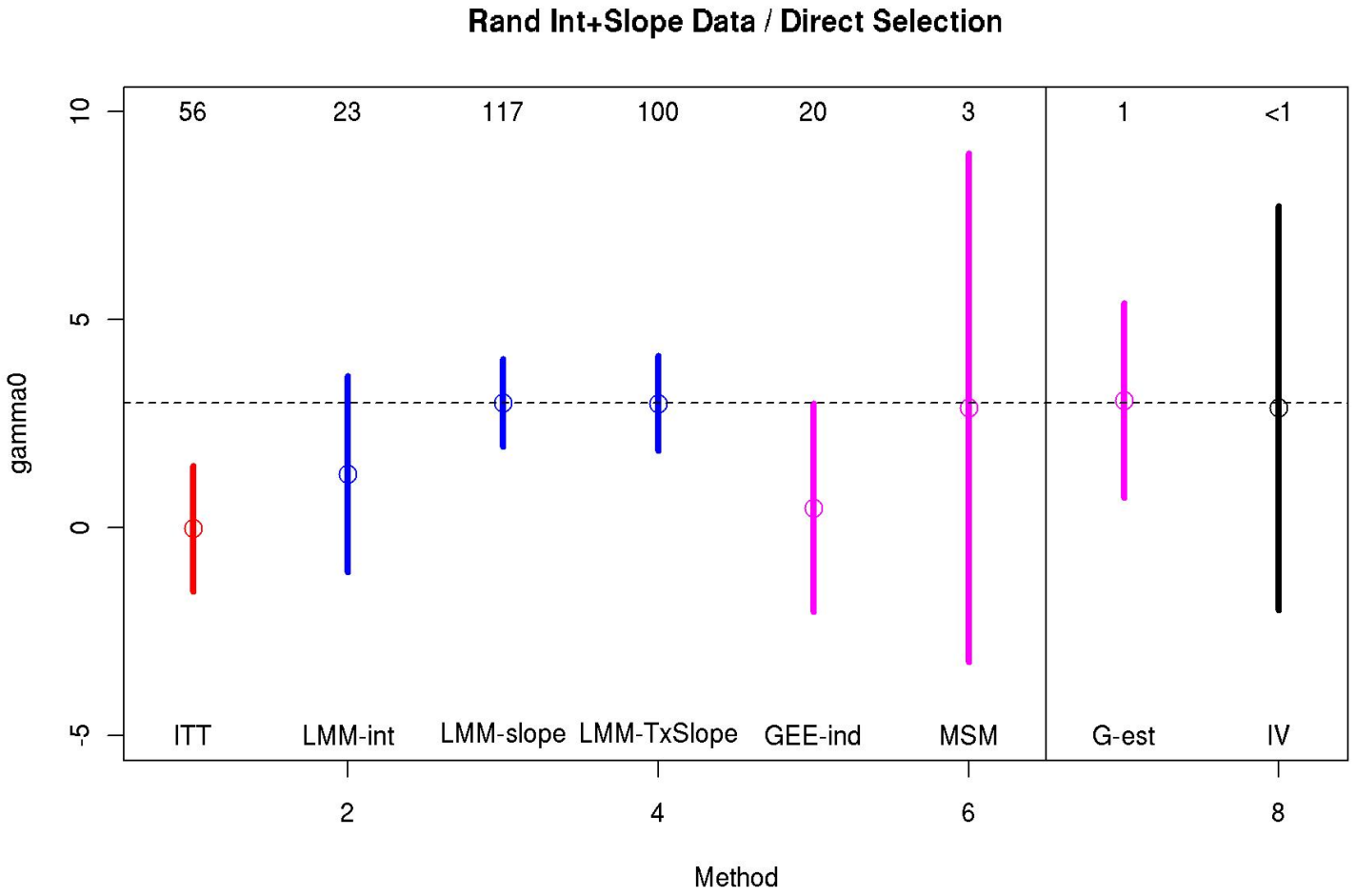
Simulations

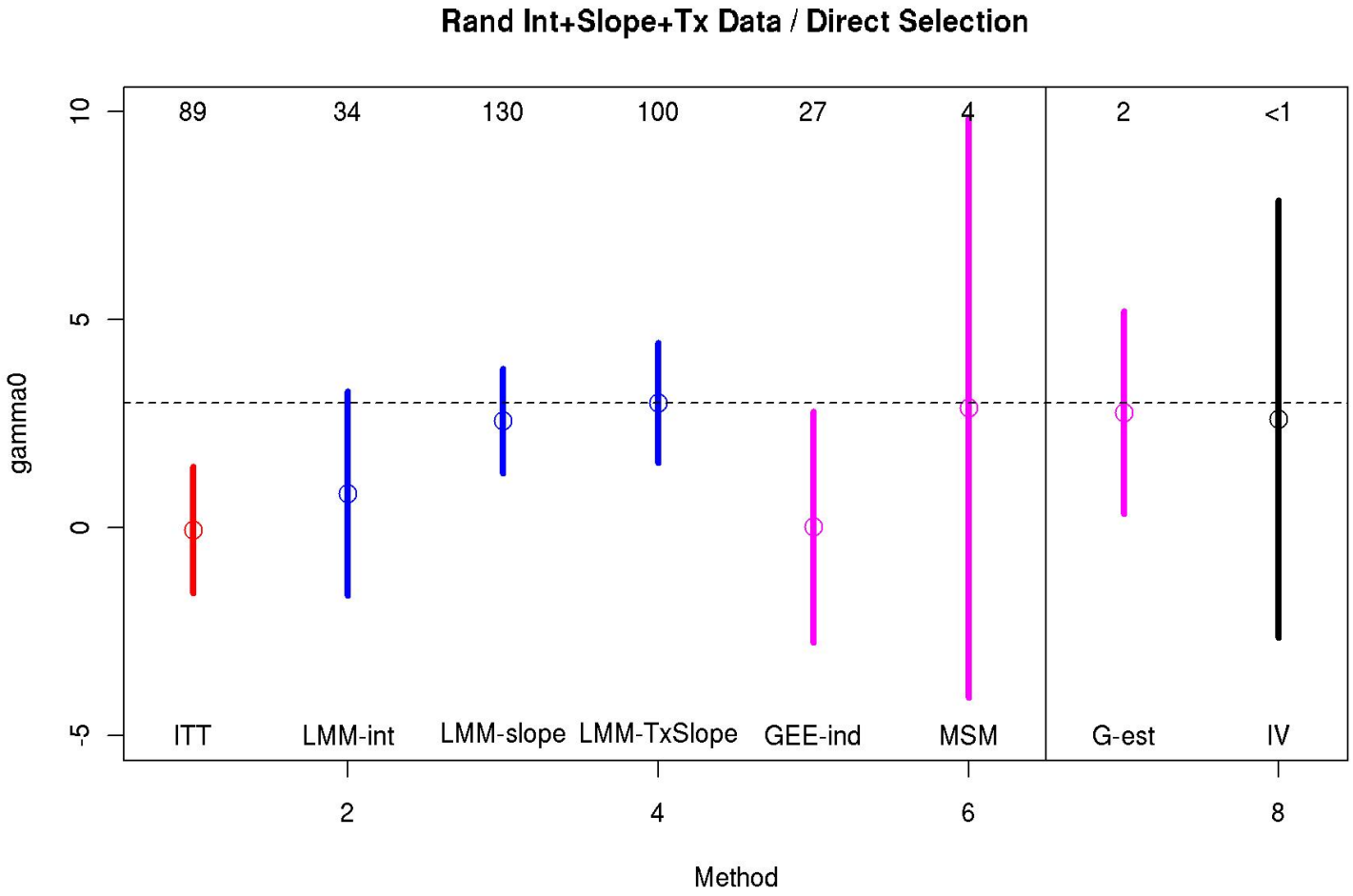
- Use a SNMM with:

$$\gamma(s,t) = \gamma_0 + \gamma_1 \cdot (t-s) = 3.0 + 0.4 \cdot (t-s)$$

- Use a Longitudinal Structural Mixed Model (LSMM)
  - ▷ Random intercept only
  - ▷ Random intercept + slope
  - ▷ Random intercept + slope + Tx
- Direct and Indirect Selection
  - ▷ logit $p_i(s) = \alpha(s, Z_i) + \eta[Y_i^O(s-1)] + \delta(b_i)$
- N=500 subjects, m=5 measurement times


Rand Int Data / Direct Selection
38
103
105
100
60
33
1
<1
10
5
0
-5
gamma0
ITT
LMM-int
LMM-slope
LMM-TxSlope
GEE-ind
MSM
G-est
IV
2
4
6
8
Method


Rand Int+Slope Data / Direct Selection
56
23
117
100
20
3
1
<1
10
5
0
-5
gamma0
ITT
LMM-int
LMM-slope
LMM-TxSlope
GEE-ind
MSM
G-est
IV
2
4
6
8
Method


Rand Int+Slope+Tx Data / Direct Selection
89
34
130
100
27
4
2
<1
10
5
0
-5
gamma0
ITT
LMM-int
LMM-slope
LMM-TxSlope
GEE-ind
MSM
G-est
IV
2
4
6
8
Method

Summary

- MLE for LSMM requires correct random effects model.
- GEE-indep biased.
- MSM and IV approx unbiased, yet high variance.
- **Q**: What justifies LSMM MLE?

Joint Likelihood Factorization

- Define:
  - $\bar{\mathbf{Y}}_i^O(t) = [Y_i^O(1), \ldots, Y_i^O(t)]$
  - $\bar{\mathbf{X}}_i(t) = [X_i(1), \ldots, X_i(t)]$ encodes S_i
- **Assume** conditional independence (no serial corr):

$$\prod_i \int \prod_t [Y_i^O(t)|\bar{\mathbf{X}}_i(t), b_i]\ [X_i(t)|\bar{\mathbf{Y}}_i^O(t-1), \bar{\mathbf{X}}_i(t-1), b_i] dF(b_i)$$

- **Assume** no indirect selection:

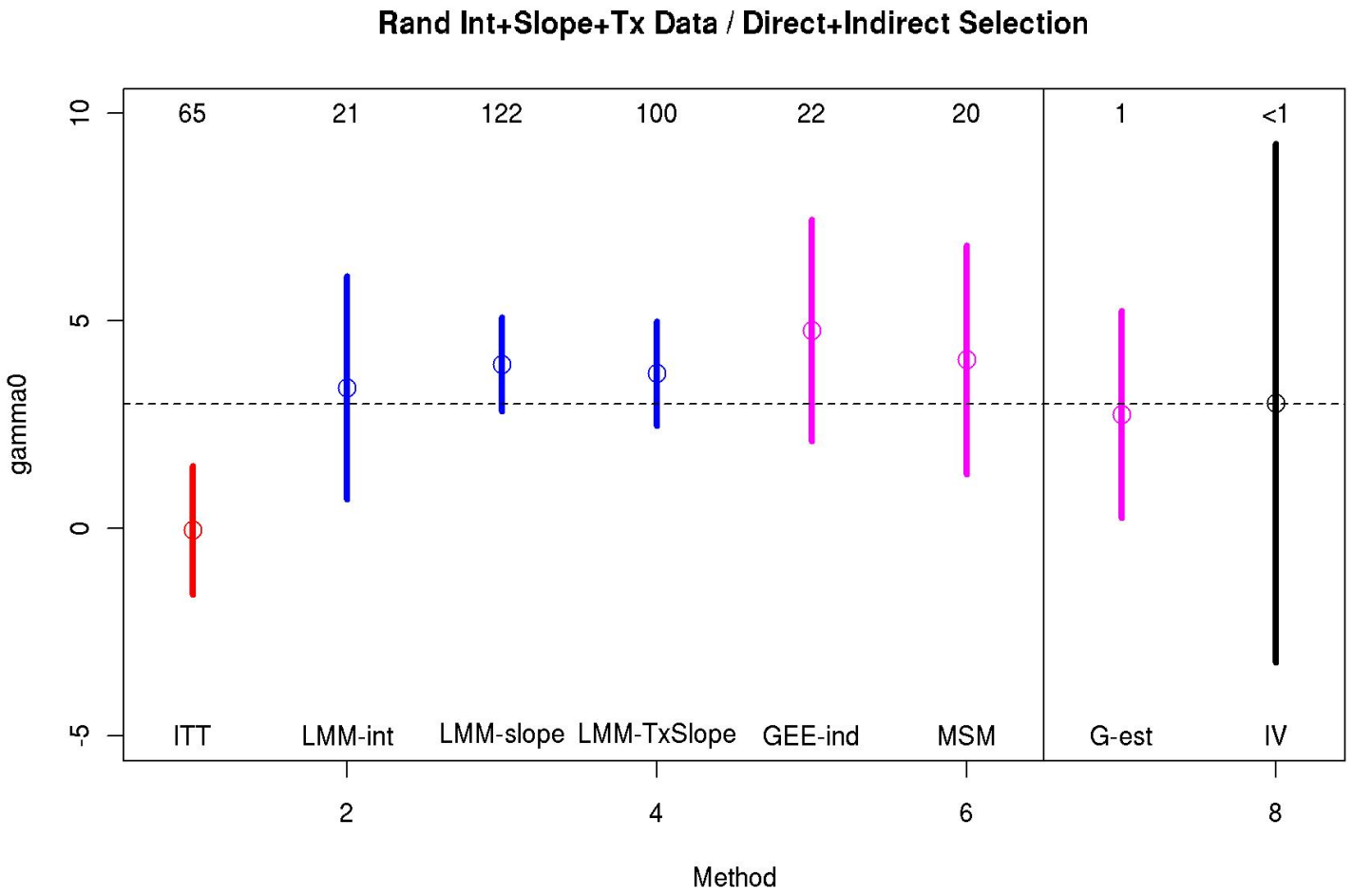
$$\prod_i \int \prod_t [Y_i^O(t)|\bar{\mathbf{X}}_i(t), b_i][X_i(t)|\bar{\mathbf{Y}}_i^O(t-1), \bar{\mathbf{X}}_i(t-1)] dF(b_i)$$

Likelihood Factorization

- Now the likelihood can be factored:

$$\prod_i \left[\prod_t [X_i(t)|\bar{\mathbf{Y}}_i^O(t-1), \bar{\mathbf{X}}_i(t-1)] \int [Y_i^O(t)|\bar{\mathbf{X}}_i(t), b_i] dF(b_i) \right]$$

  and can be maximized using typical LMM software since this is an equivalent likelihood.

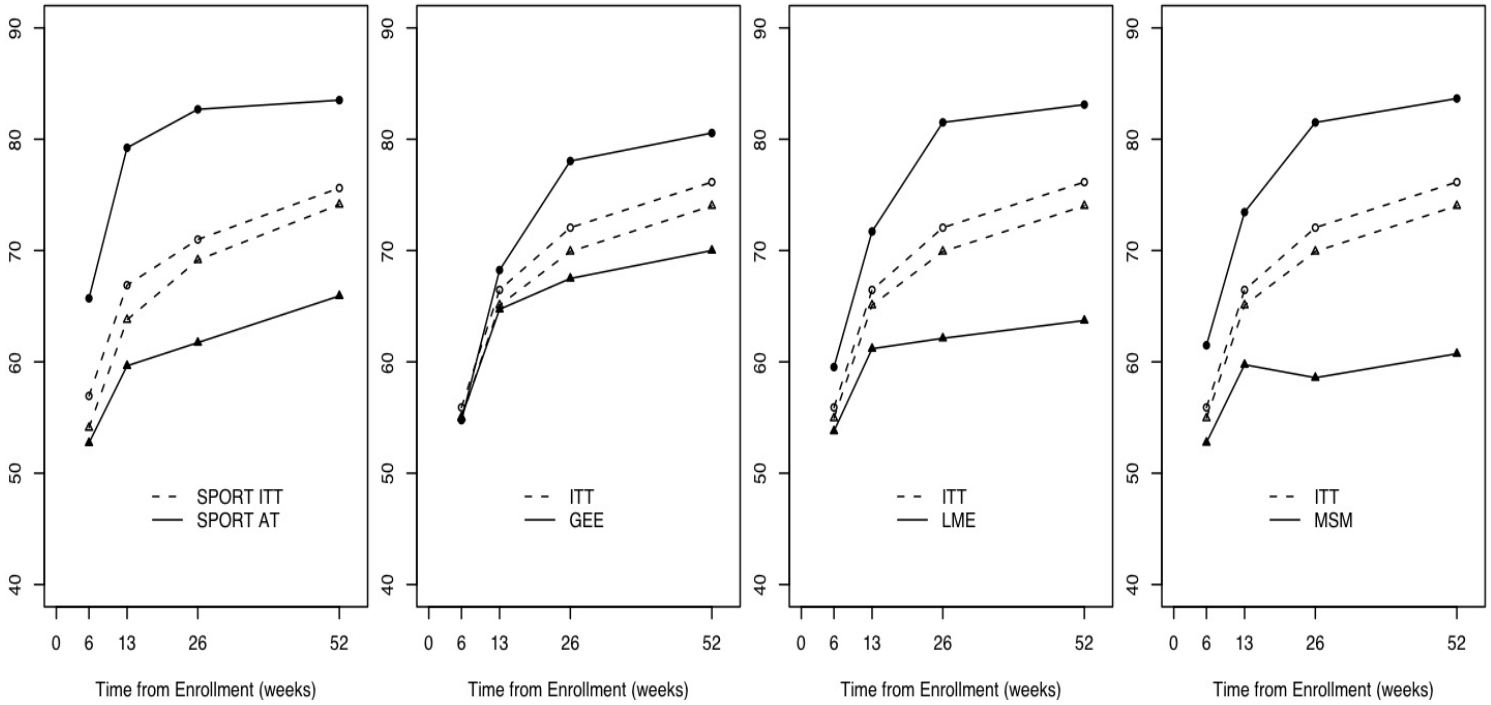

Rand Int+Slope+Tx Data / Direct+Indirect Selection
65
21
122
100
22
20
1
<1
10
5
0
-5
gamma0
ITT
LMM-int
LMM-slope
LMM-TxSlope
GEE-ind
MSM
G-est
IV
2
4
6
8
Method

Joint Likelihood (Bayes)

- Using one dataset from simulation with indirect selection we used the joint likelihood of $Y_i^O(t), S_i$ and weak priors (WinBUGS):
  - ▷ Evidence for indirect selection, where δ_5 is coefficient of b_{i2} in model for $p_i(s)$:
    $E[\widehat{\delta_5}]$=4.45 (sd[$\widehat{\delta_5}$]=1.05)
- Comparison of Methods:

	E[$\widehat{\gamma_0}$] (sd[$\widehat{\gamma_0}$])	E[$\widehat{\gamma_1}$] (sd[$\widehat{\gamma_1}$])
LSMM MLE	4.05 (0.108)	0.42 (0.0097)
Joint Likelihood	2.95 (0.199)	0.40 (0.0100)

SPORT Illustration



SPORT Illustration

Estimation	$\widehat{\gamma_0}$ (se)	$\widehat{\gamma_1}$ (se)	26-week Effect $\widehat{\gamma_0} + \widehat{\gamma_1} \cdot 26$ (se)
ITT	0.61 (2.144)	0.06 (0.0847)	**2.10** (2.082)
LME-intercept only (GEE-exchangeable)	3.06 (1.915)	0.67 (0.0968)	20.4 (1. 729)
LME-intercept + slope on time	3.61 (1.833)	0.73 (0.1004)	**22.5 (1.905)**
LME-intercept + slopes on time and X_t	1.66 (2.083)	0.68 (0.0958)	19.4 (1.972)
GEE-independence	-3.50 (2.521)	0.54 (0.1143)	10.6 (2.328)
MSM, weights: past Y level	4.48 (2.737)	0.71 (0.1431)	22.9 (3.305)
MSM, weights: past Y level and Y trend	4.05 (2.702)	0.66 (0.1398)	**21.1 (3.156)**
IV, assuming random intercepts	11.25 (31.231)	-0.14 (1.1169)	7.5 **(13.816)**
JOINT, no explicit selection	2.1 (2.173)	0.69 (0.1016)	20.0 (2.189)
JOINT, selection on b_{i0}	4.42 (2.239)	0.70 (0.1043)	22.7 (2.271)
JOINT, selection on b_{i0} and b_{i1}	-3.36 (3.071)	1.33 (0.1253)	31. 3 (3.037)
JOINT, selection on b_{i0}, b_{i1}, and b_{i2}	-6.01 (3.308)	1.25 (0. 1331)	**26.4 (4.006)**

SPORT Illustration

Estimation	$\widehat{\alpha_4}$ (sd) coeff. b_{i0}	$\widehat{\alpha_5}$ (sd) coeff. b_{i1}	$\widehat{\alpha_6}$ (sd) coeff. b_{i2}
JOINT, no explicit selection	-	-	-
JOINT, selection on b_{i0}	-0.033 (0.0048)	-	-
JOINT, selection on b_{i0} and b_{i1}	-0.041 (0.0084)	-5.41 (1.077)	-
JOINT, selection on b_{i0}, b_{i1}, and b_{i2}	-0.022 (0.0111)	-3.87 (1.236)	0.046 (0.0190)

Conclusions

- Endogenous non-compliance in surgical trials.
- Longitudinal Structural Mixed Model (LSMM).
- Selection assumptions.
- Likelihood and/or Joint Likelihood methods.
- Semi-parametric methods.

LSMM and Markers for Treatment Selection

- Motivation:
  - The **second** aim of an RCT is often to determine **who** will benefit from treatment.
  - Markers to guide treatment choice (decision)
  - Example: Carpal Tunnel / surgery / EDS and MRI
- Statistical Formulation:
  - Ability of markers to **classify**
  - Groups:

    1 : patients with `TX` $>>$ `control`

    0 : patients with `TX` $<<$ `control`

LSMM and Markers for Treatment Selection

- Typical data

subject	treatment	control	Δ
101	$Y_i(1)$	-	-
102	-	$Y_i(0)$	-

LSMM and Markers for Treatment Selection

- Desired information

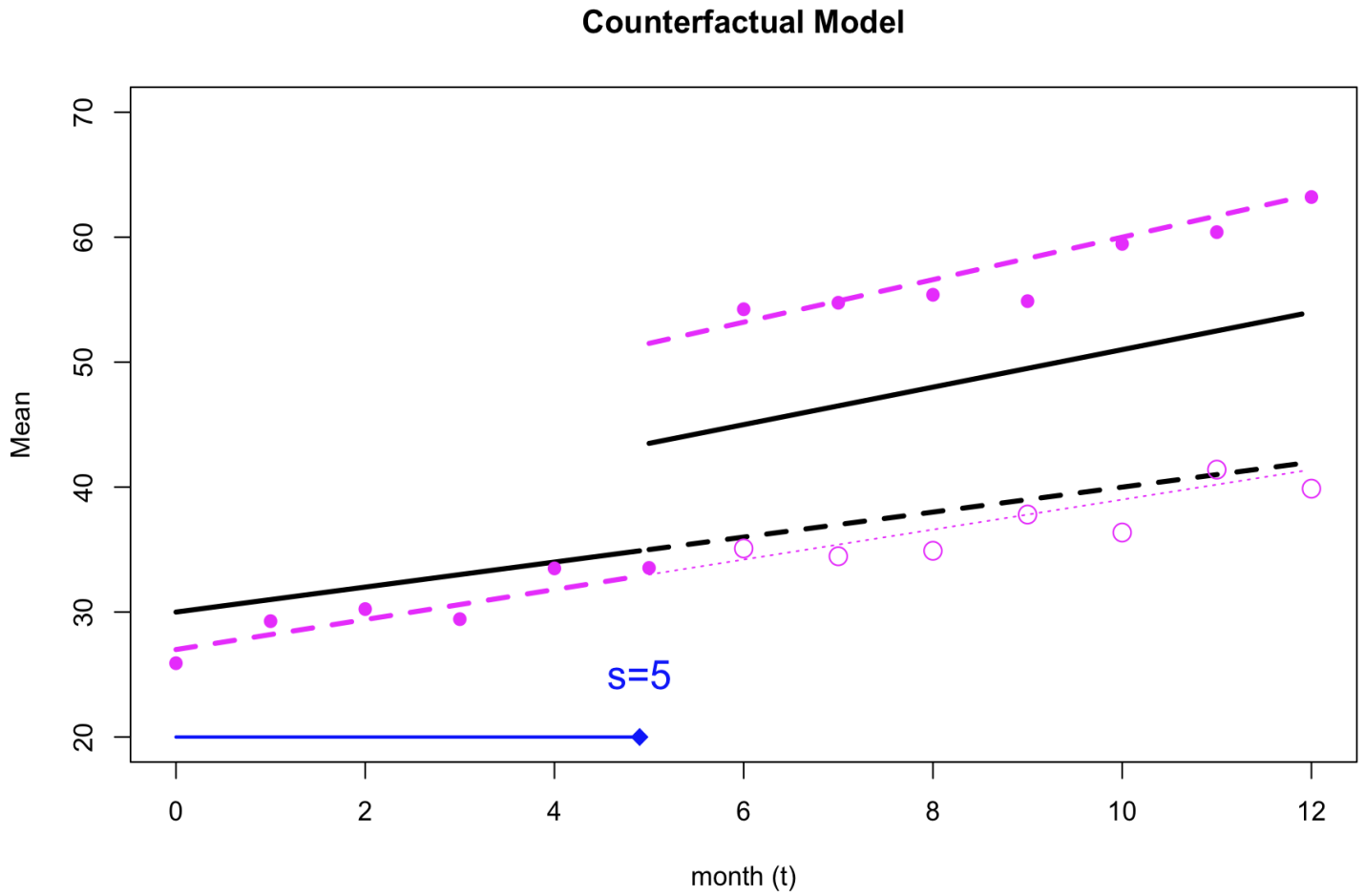
subject	treatment	control	Δ
101	$Y_i(1)$	$Y_i(0)$	Δ_i
102	$Y_i(1)$	$Y_i(0)$	Δ_i

- "Principal strata" (Frangakis and Rubin, 2002)

LSMM and Markers for Treatment Selection

- Crossover Trial

subject	time 1	time 2	Δ
101	$Y_{i1}(1)$	$Y_{i2}(0)$	$\Delta_i \approx Y_{i1}(1) - Y_{i2}(0)$
102	$Y_{i1}(0)$	$Y_{i2}(1)$	$\Delta_i \approx Y_{i2}(1) - Y_{i1}(0)$


Counterfactual Model
Mean
70
60
50
40
30
20
0
2
4
6
8
10
12
month (t)
s=5

LSMM and Prescriptive Marker

- Using the LSMM we can define individual-level treatment effects as:

$$\Delta_i(t,0) = Y_i(t,0) - Y_i(t,T^*)$$

  where $T^* > t$ indicating that subject i is non-treated at time t.

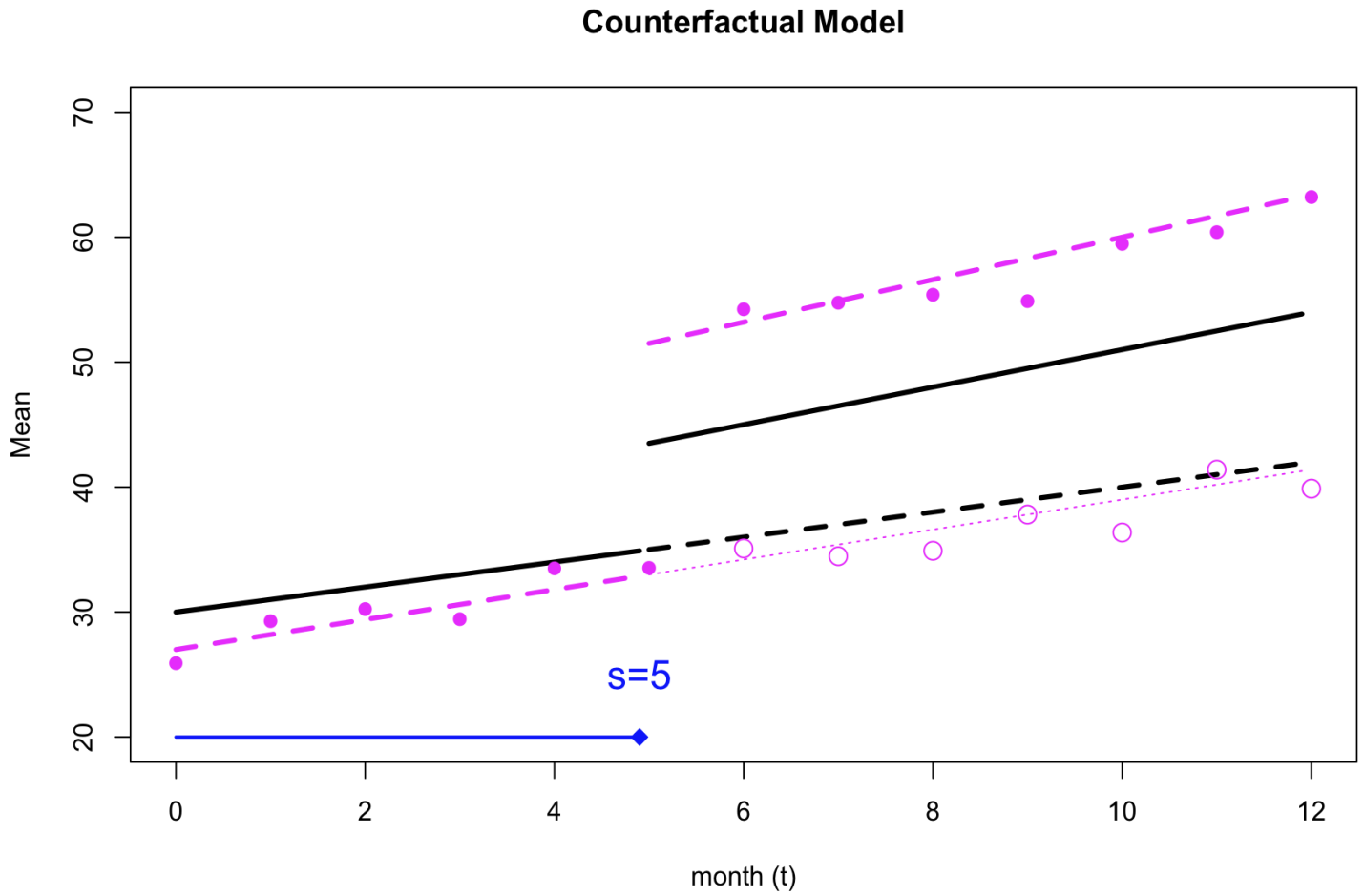
- Given data $[\bar{Y}_i(T)^O, \bar{X}_i(T)$ we will have a posterior distribution for $\Delta_i(t,0)$ since only one of $Y_i(t,0)$, $Y_i(t,T^*)$ may be observed.

LSMM and Prescriptive Marker

- Using the simple LSMM from earlier slide we would obtain:

$$\begin{aligned}\Delta_i(t,0) &= \gamma_0 + \gamma_1 \cdot t \\ &\quad + b_{i2} \\ &\quad + e_{i,1}(t) - e_{i,0}(t)\end{aligned}$$

- Note that in order to make probabilistic inference on Δ_i we would need to make error assumptions:
  - ▷ Rank preserving: $e_{i,1}(t) = e_{i,0}(t)$
  - ▷ Independence: $e_{i,1}(t) \perp e_{i,0}(t)$


Counterfactual Model
Mean
70
60
50
40
30
20
0
2
4
6
8
10
12
month (t)
s=5

LSMM and Prescriptive Marker

- Given a definition for an individual-level treatment effect we can then evaluate the ability of a marker, M_i, to classify subjects according to treatment benefit. Here we can consider:

$$\begin{aligned} \text{p-PPV} &: P[\Delta_i(t,0) > 0 \mid M_i > c] \\ \text{p-NPV} &: P[\Delta_i(t,0) \leq 0 \mid M_i \leq c] \end{aligned}$$

$$\begin{aligned} \text{p-Sensitivity} &: P[M_i > c \mid \Delta_i(t,0) > 0] \\ \text{p-Specificity} &: P[M_i \leq c \mid \Delta_i(t,0) \leq 0] \end{aligned}$$

LSMM and Prescriptive Marker

- In order to make inference a joint model for the outcome(s) and the marker is needed. Use $[Y_i \mid M_i] \cdot [M_i]$:

$$\begin{aligned} E[Y_i(t,s) \mid M_i] &= \beta(t, M_i) + \gamma(t, s, M_i) \cdot 1(t > s) \\ &\text{e.g.} \quad \gamma_0 + \gamma_1 \cdot (t - s) + \gamma_2 \cdot M_i \\ b_i, e_i &\sim \mathcal{N} \\ M_i &\sim F_M \end{aligned}$$

- Either a parametric or non-parametric model can be assumed for M_i.

LSMM and Prescriptive Marker

- Our current work is on inference for the classification error rates p-Sensitivity and p-Specificity. Here the development uses:

$$\begin{aligned} \text{p-Sensitivity} &= P[M_i > c \mid \Delta_i > 0] \\ &= \frac{\int_c^{\infty} P[\Delta_i > 0 \mid M_i = m] \cdot P[M_i = m]}{\int_{-\infty}^{\infty} P[\Delta_i > 0 \mid M_i = m] \cdot P[M_i = m]} \end{aligned}$$

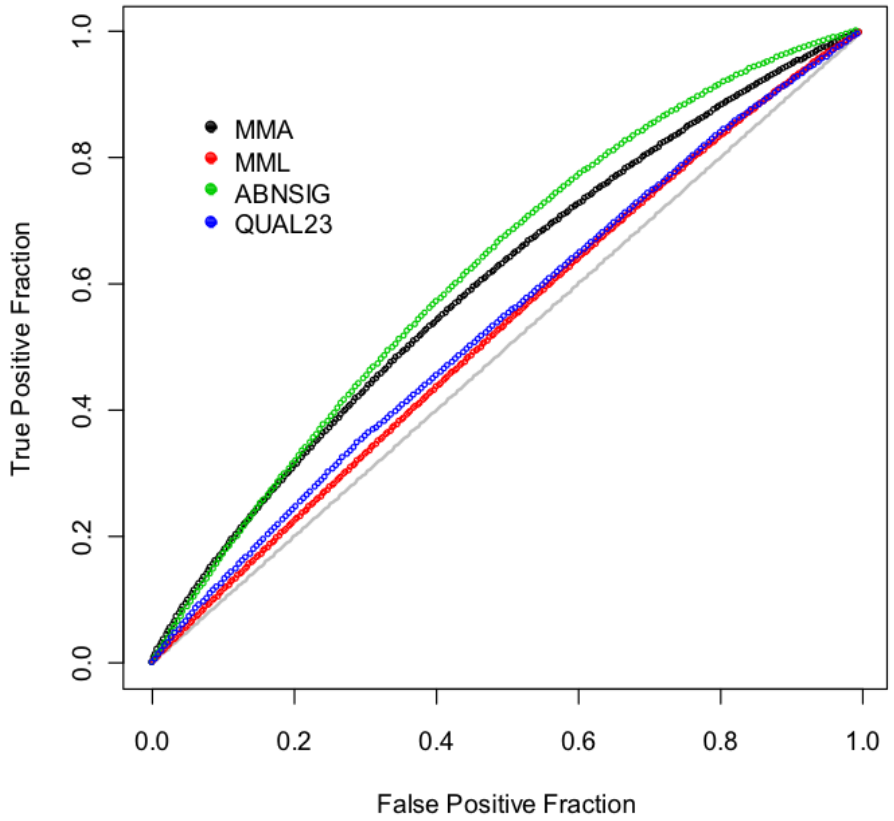
- Prescriptive ROC curves can be obtained by varying the value of c.
- Other thresholds for benefit: $\Delta_i > d$.

Surgery for Carpal Tunnel Syndrome

- Carpal Tunnel | Jarvik et al. (2009) *Lancet*
- Subjects with mild-to-moderate carpal tunnel syndrome (CTS).
- RCT subjects (N=116) and observational arm (N=207)
- Hand functional status measured at t=0, 3, 6, 9, 12 months

CTS: Markers to Indicate Surgery

- Jarvik et al. (2008) and (2009) identified variables that were associated with treatment benefit.
- Electrodiagnostic (EDS) and MRI markers
  - EDS: Median motor latency (MML)
  - EDS: Median motor amplitude (MMA)
  - MRI: Abnormal signal length (ABNSIG)
  - MRI: Bowing of flexor retinaculum (QUAL23)
- **Q**: How do the markers compare in their ability to accurately target treatment to those subjects who will benefit?


MMA
MML
ABNSIG
QUAL23
True Positive Fraction
False Positive Fraction
0.0
0.2
0.4
0.6
0.8
1.0

LSMM: Summary

- Structural models specify parameters of interest.
- Alternative methods of estimation for treatment effects.
- Extension to evaluation of markers.
- Evaluation of prescriptive classification error rates.
- Collaboration with Colleen Sitlani (UW).

Our Manuscripts

- Sitlani, Heagerty, Tosteson, Blood: Longitudinal structural mixed models for the analysis of surgical trials with non-compliance. (submitted)
- Sitlani, Heagerty, Comstock: Longitudinal structural mixed models as tools for characterizing the accuracy of markers used to select treatment. (manuscript)