Virtual Twins method for estimating long-term treatment effects from a short-term clinical trial with an active-arm extension

- Background, trial extension designs
- Alternative analysis methods
- Virtual twins method
- Application to FIT/FLEX
- Model misspecification
- Simulation study
- Conclusions

Two disclosures: Chuck McCulloch's idea, salary support from Amgen

Background

- Treatment shown to work well in short-term placebo-controlled trial
- But would it work as well, and benefits outweigh adverse effects, in long-term use?
- Long-term placebo-controlled trials infeasible for ethical, practical reasons
- Evidence restricted to short-term placebo-controlled trials with various longer-term extensions

Extension designs - I

trial		extension
active	$\rightarrow$	none
placebo	$\rightarrow$	none

- IBIS, BCPT (tamoxifen for breast cancer prevention): blinded follow-up, no one remained on active treatment
- Captures long-term effects of short-term treatment, but not effects of long-term treatment

Extension designs - II

trial		extension
active	$\rightarrow$	active
placebo	$\rightarrow$	none

- MARS (lovastatin for atherosclerosis): double-blind extension, 58% participation; stopped by DSMB
- HERS (hormone therapy for secondary prevention of CHD):
  - in trial, no overall treatment effect – early harm, late benefit
  - participants unblinded, encouraged not to change treatment
  - 93% participation, little crossover in 2.8 year extension; no long-term benefit

Extension designs - III

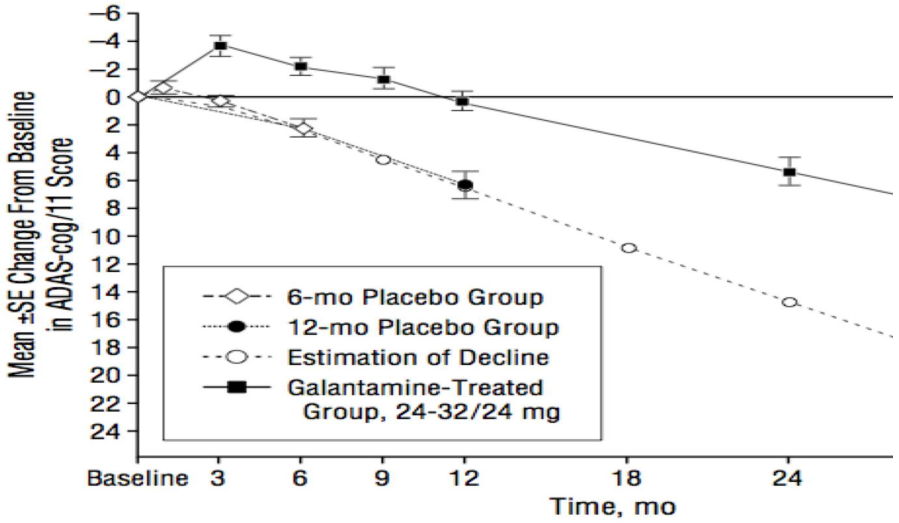
trial		extension
active	$\rightarrow$	active
placebo	$\rightarrow$	active

- Most common design
- Analyses focus on within-group changes from baseline in active$\rightarrow$active group
- Examples: bisphosphonates (Bone *et al.*, *NEJM*; Tonino *et al.*, *JCEM*), clozapine in late Parkinson's, galantamine for dementia, possibly iPrEx pre-exposure prophylaxis trial

Virtual Twins pre-cursor

- Galantamine for treatment of Alzheimer's
- Placebo group crossed over to active for extension period
- Long-term placebo response projected forward, using baseline scores of placebo group and published prediction equations based on historical data

Virtual Twins pre-cursor



Extension designs - IV

trial		extension
active	$\rightarrow$	active/placebo
placebo	$\rightarrow$	discontinued

- FIT/FLEX
- Randomized active/placebo comparison in extension shows whether is it better to continue on treatment than to stop
  - but not whether it is better to have started in the first place
- Active$\rightarrow$placebo group not informative for long-term effects

Extension designs - summing up

- Most include long-term active treatment
- Almost none include long-term placebo
  - can't control for long-term placebo, secular effects
- Extension usually requires re-consent
  - can entail considerable dropout, selection bias

Naive estimators of long-term treatment effects

- Assume short-term effect holds long-term
- Assume short-term effect holds long-term, provided trial and extension outcome rates equivalent in active arm
- FIT/FLEX design: estimate late treatment effect by active/placebo comparison in extension, combine with short-term effect from trial
- Project placebo forward: estimate late treatment effect by comparing active in extension to placebo in trial, combine with short-term effect from trial

Virtual Twins estimator

- *For each active-treatment volunteer in the extension, model expected outcomes for a virtual twin with the same prognostic covariates, under counterfactual assignment to placebo*
- Parameters of outcome models estimated using data from placebo group
- Expected outcomes for each twin calculated using those parameter estimates, covariate values for volunteer
- Treatment effects estimated by average difference (or ratio) of observed outcomes for volunteers, expected outcomes for virtual twins
- Bootstrap used to calculate confidence intervals

Controlling selection effects

- Re-consent for extension opens door to selection bias
- Controlled by computing expected outcomes for twins using prognostic covariates for volunteers
- Assumes that measured covariates adequately capture selection effects
- Inference restricted to volunteer group

Controlling secular effects

- Placebo outcome rates might change during extension
- Controlled by updating covariates for twins:
  - use observed end-of-trial values for volunteers if unaffected by treatment
  - otherwise simulate end-of-trial values:
    1. fit models for end-of-trial values in placebo group
    2. simulate values for twins using model parameter estimates, baseline covariates for volunteers

Counterfactual framework of extension data

Group	Volunteer for Extension	Proportion of Group	Responses	
			Trial	Extension
Treatment	Yes	$p_{v,T}$	$Y^{v}_{1,T}$	$Y^{v}_{2,T}$
	No	$1-p_{v,T}$	$Y^{nv}_{1,T}$	$Y^{nv}_{2,T}$
Placebo	Yes	$p_{v,P}$	$Y^{v}_{1,P}$	$Y^{c}_{2,P}$
	No	$1-p_{v,P}$	$Y^{nv}_{1,P}$	$Y^{nv}_{2,P}$

- highlights show what we don't observe in FIT/Flex design

If we had complete data ...

- Effect of treatment during trial is

$$[p_{v,T}\mu^{v}_{1,T} + (1 - p_{v,T})\mu^{nv}_{1,T}] - [p_{v,P}\mu^{v}_{1,P} + (1 - p_{v,P})\mu^{nv}_{1,P}],$$

  where $\mu^{v}_{1,T} = \mathrm{E}[Y^{v}_{1,T}]$, and so on

- Treatment effect during the extension is

$$[p_{v,T}\mu^{v}_{2,T} + (1 - p_{v,T})\mu^{nv}_{2,T}] - [p_{v,P}\mu^{v}_{2,P} + (1 - p_{v,P})\mu^{nv}_{2,P}]$$

- Long-term effects of treatment estimated by (weighted) average of trial and extension differences

In absence of complete data ...

- Since extension responses for non-volunteers are never observed, we can at best estimate treatment effects in volunteers: $\mu^v_{1,T} - \mu^v_{1,P}$ and $\mu^v_{2,T} - \mu^v_{2,P}$
- Virtual Twins method estimates the expected counterfactual placebo responses of each observed active-treatment volunteer
- Two potential problems to resolve:
  - Selection effects: $\mu^v_{i,P} \neq \mu^{nv}_{i,P}$?
  - Secular, cohort effects: $\mu^v_{2,P} \neq \mu^v_{1,P}$?

Some notation

- $\mathbf{X_1^i}$: vector of prognostic baseline covariates that are fixed or not affected by treatment
- $\mathbf{X_1^d}$: other prognostic baseline covariates that are subsequently affected by treatment
- $\mathbf{X_2^i}$ and $\mathbf{X_2^d}$: updated values of baseline covariates, observed or counterfactual, at start of extension
- $\mathbf{Z}$: additional baseline covariates that influence $Y_{2,P}^v$ only through their effects on $\mathbf{X_2^d}$

Model for virtual twin responses during the trial

- Assume placebo responses during trial arise from GLM with conditional mean

$$\mathrm{E}[Y_{1,P}|\mathbf{X_1^i}, \mathbf{X_1^d}] = \mathbf{g^{-1}}[\beta_\mathbf{0} + \boldsymbol{\beta}_\mathbf{1}\mathbf{X_1^i} + \boldsymbol{\beta}_\mathbf{2}\mathbf{X_1^d}] \qquad (1)$$

- Model (1) assumed to hold for all placebo participants: i.e., $\mathbf{X_1^i}$ and $\mathbf{X_1^d}$ capture any dependence of response on volunteering for the extension (selection effects)

- In type III design, $\boldsymbol{\beta}$ could be estimated using using data for placebo volunteers only

Complete data model for virtual twin responses during the extension

- Assume that during the extension, counterfactual placebo responses arise from the same GLM as (1), but with conditional mean

$$\mathrm{E}[Y^v_{2,P}|\mathbf{X^i_2}, \mathbf{X^d_2}, \mathbf{Z}] = \mathbf{g^{-1}}[\beta_\mathbf{0} + \boldsymbol{\beta_1}\mathbf{X^i_2} + \boldsymbol{\beta_2}\mathbf{X^d_2}] \quad (2)$$

- Link function, parameters shared by (1) and (2)
- Equivalently: differences in conditional means of $Y^v_{1,P}$ and $Y^v_{2,P}$ due to secular, cohort effects completely captured by changes in $\mathbf{X^i}$ and $\mathbf{X^d}$

Prognostic variables at start of extension

- Use observed values of $\mathbf{X_2^i}$ for volunteers in evaluating (2) for virtual twins (values are not affected by treatment)
- Model counterfactual values of $\mathbf{X_2^d}$, assuming each element arises from GLM with conditional mean

$$\mathrm{E}[X_{2j}^d|\mathbf{Z_j}] = \mathbf{h_j^{-1}}[\boldsymbol{\gamma_{0j}} + \boldsymbol{\gamma_{1j}}\mathbf{Z_j}],\ \mathbf{j} = \mathbf{1}, \ldots, \mathbf{J}. \qquad (3)$$

- $\boldsymbol{\gamma}$ estimated using placebo data during trial
- $\mathbf{Z}$ may include elements of $\mathbf{X_1^d}$ (i.e., baseline values of the treatment-affected covariates used as predictors)

Observed data model for virtual twin responses during the extension

- Combining (2) and (3), we obtain

$$\mathrm{E}[Y^{v}_{2,P}|\mathbf{X^{i}_{2}},\mathbf{Z}] = \int \cdots \int \mathbf{g^{-1}}[\beta_{\mathbf{0}}+\boldsymbol{\beta}_{\mathbf{1}}\mathbf{X^{i}_{2}}+\boldsymbol{\beta}_{\mathbf{2}}\mathbf{X^{d}_{2}}]\mathbf{dF_{1}}\cdots\mathbf{dF_{J}}, \tag{4}$$

  where $\mathbf{dF_{j}}$ is conditional density of $\mathbf{X^{d}_{2j}}$ given $\mathbf{Z_{j}}$, consistent with (3)

- (4) estimated using Monte Carlo integration:
  - sample $\mathbf{X^{d}_{2}}$ from (3) using $\hat{\boldsymbol{\gamma}}$
  - compute $\mathrm{E}[Y^{v}_{2,P}|\mathbf{X^{i}_{2}},\mathbf{X^{d}_{2}},\mathbf{Z}]$ using (2) and $\hat{\boldsymbol{\beta}}$
  - repeat and average the results

Effects of treatment among volunteers

- If assumptions hold, estimate treatment effect during trial by averaging

$$Y^v_{1,T} - \mathrm{E}[Y^v_{1,P}|\mathbf{X^i_1}, \mathbf{X^d_1}]$$

  over the volunteer sample

- Similarly, estimate treatment effect during extension by averaging

$$Y^v_{2,T} - \mathrm{E}[Y^v_{2,P}|\mathbf{X^i_2}, \mathbf{Z}]$$

  again over the volunteers

- Estimate long-term effect of treatment by weighted average of trial and extension effects

Bootstrap CIs

- Variability arises from
  - sampling of volunteers
  - sampling of placebo participants used to estimate model parameters
- Solution: resample with replacement from volunteers and placebo group, re-run procedure on each bootstrap sample
- Compute confidence bounds as percentiles of bootstrap effect estimates
- Compute point estimate as mean of effect estimates, averaging over simulations of end-of-trial covariates

Fracture Intervention Trial (FIT)

- Two large RCTs of Alendronate (ALN) for prevention of fractures
- Vertebral fracture trial:
  - 2027 post-menopausal women with existing vertebral fracture (VFx)
  - randomized 1-1 to ALN or placebo
  - 2.9-year average follow-up
  - primary endpoint: new morphometric VFx
  - results (N=1946): RR 0.53, 95% CI 0.41-0.68

FIT

- Clinical fracture trial:
  - 4432 post-menopausal women with low BMD
  - randomized 1-1 to ALN or placebo
  - 4.2-year average follow-up
  - primary endpoint: clinical fracture
  - results:
    - overall (N=4272): RR 0.86, 95% CI 0.73-1.01
    - T-score < -2.5: RR 0.64, 95% CI 0.50-0.82

FLEX - the extension

- At end of FIT, one year of ALN offered to all participants; outcomes not ascertained in this interval
- After interim open-label period (average 1.9 years), new trial of 3 additional years of ALN or placebo
- Eligibility for FLEX:
  - assigned to ALN in FIT
  - ≥ 3 years of ALN during trial and interim peirod
  - T-score > -3.5, BMD $>$ FIT baseline
- Of 3236 assigned to ALN in FIT, 1099 randomized 3-2 in FLEX to ALN or placebo
- 662/1099 volunteers assigned to ALN included in analysis

FIT/FLEX Virtual Twins analysis

- Outcome: number of nonspine clinical fractures (NSFx)
  - Poisson, not-overdispersed
  - log-transformed volunteer person-years used as offsets
- Covariates unaffected by treatment: age, BMI, smoking
- Covariates potentially affected by treatment: history of NSFx, VFx, BMD
- NSFx, VFx, BMD modeled using Poisson, logistic, linear models

Results for FIT/FLEX

	Rates per 100 person-years (95% CI)		
	Trial	Extension	Overall
Volunteers	3.6 (2.7-4.6)	5.1 (4.2-6.0)	4.4 (3.7-5.2)
Virtual twins	4.3 (3.8-4.7)	5.1 (4.4-5.8)	4.7 (4.2-5.3)
Rate ratio	0.85 (0.61-1.10)	1.00 (0.80-1.22)	0.94 (0.77-1.11)

Naive estimates of effect of long-term treatment for FIT/FLEX

1. Assume short-term results hold long-term: RR 0.83 (0.73-0.96)
2. Conditionally assume short-term results hold long-term: fracture rates in active arm increased from 3.6 to 5.1 per 100 p-y ($p < 0.0001$)
3. Use FLEX ITT result to estimate late treatment effect: 0.99 (0.77-1.26). Overall ITT FIT/FLEX RR: 0.87 (0.76-0.996)
4. Project placebo results: overall RR 0.94 (0.82-1.09)

Assumption checking: selection effects

- Outcome models assumed to hold for all placebo participants: covariates capture any dependence of response on volunteering for the extension
- In FIT/FLEX, placebo participants were not asked to participate in extension, so we can't estimate parameters using data for placebo volunteers only
- In FIT, NSFx and VFx rates were lower, end of study BMD higher, among FLEX volunteers, compared to other ALN particpants
- "Offset model" didn't clearly help in simulations

Assumption checking: secular effects

- Overall placebo rates in trial could mask an increasing trend ("healthy volunteer effects"), so calculated twin rates in extension would be biased low
- If interim outcomes during trial available, check for trend independent of time-dependent prognostic covariates
- Parameters for calculating expected extension rates for twins could be estimated omitting early trial data, but extrapolation would be problematic
- "Delay model" works for healthy volunteer effect in trial, fails with further changes in rates in extension
- Also: consider plausibility of other secular effects during extension

Simulation studies

- Assessed bias of RR estimate, CI coverage and width, relative efficiency
- Selection bias: correlated random effects used to link BMD change, fracture risk, volunteering for extension
- Secular bias: placebo event rates increased in second half of trial and extension, beyond what is predicted by baseline and end-of-trial covariates
- Relative efficiency: mean squared error compared to long-term placebo-controlled trial

Simulation study of selection effects

	Percent	Trial		Extension		Overall	
Scenario	Volunteering	Bias	Cov	Bias	Cov	Bias	Cov
True model	35%	-0.2	94.8	-0.5	97.6	0.2	96.4
Selection	35%	-3.5	92.4	-3.6	93.6	-3.0	93.6
Selection	85%	-0.8	96.8	0.2	93.2	-0.1	93.2

Bias - Percent bias of rate-ratio estimate

Cov - Coverage of 95% confidence interval for rate-ratio

Simulation study of secular effects

	Secular	Trial		Extension		Overall	
Model	Increase	Bias	Cov	Bias	Cov	Bias	Cov
Standard	Trial only	0.8	93.2	28.4	49.2	17.2	70.0
	Both	0.7	95.6	62.4	0.8	37.2	14.4
Delay	Trial only	0.7	95.2	1.0	94.8	0.8	94.0
	Both	0.8	93.6	29.9	64.0	9.3	70.0

Bias - Percent bias of rate-ratio estimate

Cov - Coverage of 95% confidence interval for rate-ratio

Delay model uses second half of trial data to estimate parameters used in calculating expected twin outcomes in extension

Simulation study of relative efficiency

		Trial		Extension		Overall	
Scenario	Model	MSE	RE	MSE	RE	MSE	RE
True model	Full data	0.022		0.018		0.010	
	Twins	0.018	1.21	0.011	1.56	0.009	1.05
Selection (35%)	Full data	0.028		0.012		0.009	
	Twins	0.017	1.62	0.014	0.89	0.010	0.92
Selection (85%)	Full data	0.009		0.007		0.004	
	Twins	0.008	1.19	0.008	0.82	0.006	0.68

MSE - Mean Squared Error of rate-ratio estimate

RE - Relative Efficiency

Conclusions

- A method for estimating effects of long-term treatment from extension studies when the only placebo data is short-term
- Makes fewer and less onerous assumptions than naive methods; assumptions can be partially checked
- Simulations suggest selection bias benign, secular effects might cause trouble; relative efficiency surprisingly good
- Our view: preferable to alternative methods across the board
- Application to HIV/AIDS: possible extension of iPrEx trial of pre-exposure prophylaxis. Others?