

Innovations in the Design and Analysis of Trials Using Electronic Health Care Data

Ross Prentice
Fred Hutchinson Cancer Research Center
University of Washington
October 24, 2016

Today's Presentations

- EHR/EMR for research, including observational research (big data, precision medicine; data quality/supplemental data)
- Embedding (prevention or treatment) clinical trials into a context where EHR data are available (Cook, Hughes, Rosenblum)
- Embedding clinical trials into a mHealth context (Klasnja, Murphy)

Initial Comments

- Observational research is a crucial component of the public health and biomedical research agendas.
- Observational research is extremely challenging, especially for the evaluations of treatments/interventions (e.g., menopausal hormone therapy) and especially for difficult to measure exposures (e.g., diet and physical activity).
- The interface between observational studies and clinical trials is an under-developed component of the public health/chronic disease prevention research arena.
- The preventive intervention development enterprise is in need of innovation and additional attention.

Embedding Treatment or Prevention Clinical Trials into a Context Where EHR Data are Available

An excellent research emphasis given the efficiencies that may be possible, compared to traditional clinical trial contexts, for recruitment, participant monitoring, data collection...

Andrea Cook: Pragmatic Clinical Trial Challenges: Lessons Learned from the NIH Collaboratory Biostatistics and Design Core

- Pragmatic vs. Explanatory Trials
- Demonstration projects – patient care, but also disease screening enhancement, suicide prevention
- Trial design- individual versus cluster, and related sample size and power issues
- Choice of clusters, range of data analysis types (marginal methods; mixed models/frailty or copula models)
- Crossover designs (carry-over effects); stepped wedge design (power); analytic approaches (multivariate failure time methods needed); randomization approaches; outcome data quality
- Questions that presumably cannot be answered using EHR

Embedding Treatment or Prevention Clinical Trials into a Context Where EHR Data are Available

Jim Hughes: Design and Analysis of Stepped Wedge Trials

- Cluster randomized trials and their possible designs
- Advantages and disadvantages of stepped wedge designs
- Cross-sectional or cohort sampling
- Data analysis options:
 - (i) Analyses in which each cluster acts as its own control
The fact that temporal variations in outcome rates need to be controlled by regression modeling may be a major disadvantage. Also '(likely) fewer clusters needed'?
 - (ii) Vertical analysis
Can be considered as marginal methods with time-varying intervention group indicator: Avoids the need for adjustment for temporal variations, but efficiency may be poor? Can efficiency be improved by some non-uniform intervention cluster timing designs?
- Data analysis methods for censored failure time outcomes?

Embedding Treatment or Prevention Clinical Trials into a Context Where EHR Data are Available

Michael Rosenblum: Optimizing Adaptive Enrichment Designs, and Challenges in Using Data to Construct Realistic Simulations to Evaluate Design Performance

- Adaptive designs, potential benefits and risks
- Based on interim analyses, change study design (sample size, eligibility/exclusionary criteria, follow-up duration, intervention dose, randomization fraction, ...)
- SMART designs – If participant ‘fails’ on initial treatment, then may randomize to another. Only short-term outcomes can be entertained.
- PCORI-funded project toward designs that can create stronger evidence about subpopulation benefits and harms (time-to-event outcomes)
 - Two subpopulations, control error rates, satisfies specified power criteria in subpopulations and overall, optimizes sample size/duration
- Data generation for simulation evaluation of RCT properties
 - Resample from real datasets (with/without replacement)
 - Hemorrhagic stroke (<10ml vs. ≥10ml)
 - MCI to Alzheimer’s (APOE₄ carrier status)

Incorporating Clinical Trials into the mHealth Context

Padja Klasnja: Micro-randomized Trials: The Design for Optimizing Just-in-Time Interventions

Susan Murphy: Some Data Analytics for Developing Just-in-Time Adaptive Interventions in Mobile Health

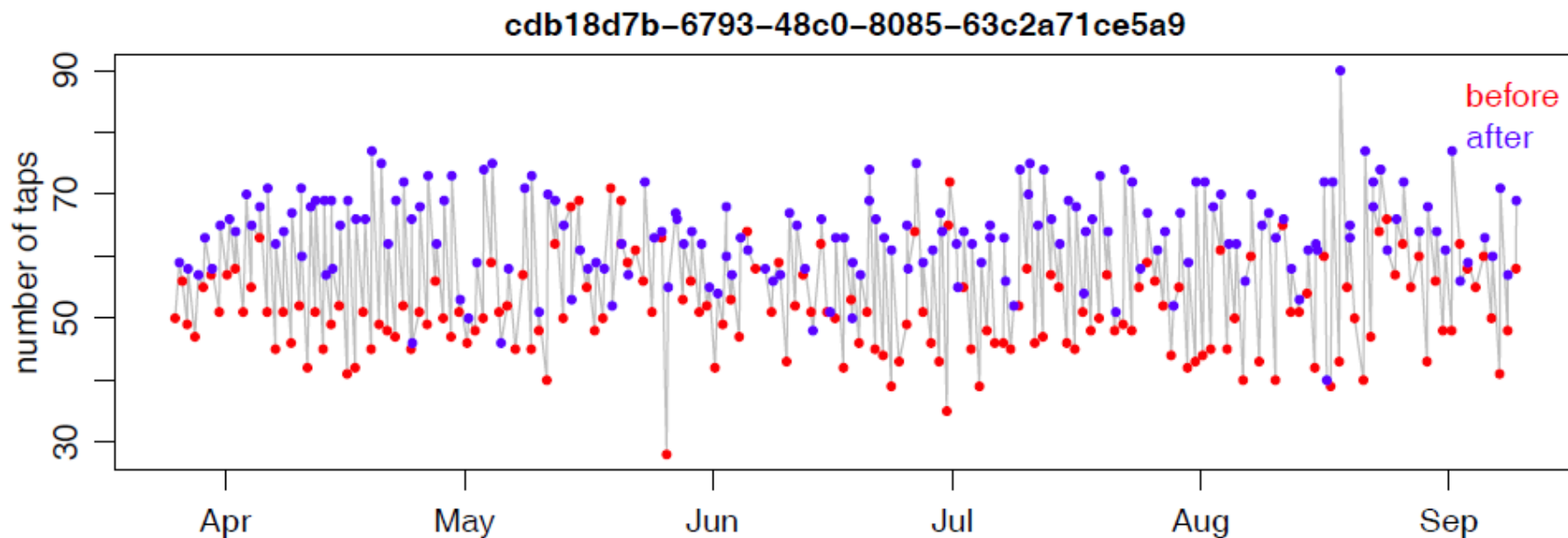
Potential to study a broad range of interventions, for a broad range of outcomes, for a broad range of populations worldwide

e.g., Develop and evaluate interventions to achieve a desired behavior, such as daily walking goals in Heartsteps

e.g., Develop and evaluate interventions to reduce or manage chronic disease symptoms, such as Parkinson's disease symptom management in mPower as part of Apple's Research Kit mobile platform. Over 9000 participants enrolled in first 6 months with outcome of smartphone screen taps over 20 seconds in relation to whether tests are performed before or after levodopa medication.

Some related analytics in submitted paper: Chaibub Neto E, Prentice RL, Bot BM, Kellen M, Friend SH, Tristen AD, 2016.

Example of data collected by the mPower study. The data is color coded according to whether the number of taps was recorded before (red dots) or after (blue dots) the participant took medication



Micro-Randomized Trial Design

(Peja)

Design goals:

- Effective long-term interventions
- Adaptive to individuals' circumstances
- Delivered at the appropriate times and places (only when participant is 'available')

Evaluation goals:

- Effectiveness of intervention (components)
- Delivery times, contexts, user burden
- Dependence of effectiveness on context, individual characteristics

Evaluation summaries:

- Time-varying intervention effects
- Contextual modifiers
- Can only study proximal outcomes (that may be mediators of longer-term outcomes)

Micro-Randomized Trial Design

(Peja)...

Comment: Trials of this type may lead to the identification of behavior change interventions of much greater effects than has been the case with traditional RCTs.

Question: Can an intervention developed using micro-randomized trials, serve as the intervention for a subsequent chronic disease prevention trial (e.g., in the dietary or physical activity epidemiology area)?

Micro-Randomized Trial Analytics

(Susan)

Individual-level data: $O_1, A_1, Y_2, \dots, O_t, A_t, Y_{t+1}$

O^t – observations at t^{th} decision time (includes individual availability)

A^t – intervention at t^{th} decision time (including whether to provide intervention)

Y^{t+1} – proximal outcome following A^t (step count over 30 minutes)

Models:

$$Y_{t+1} = a_0 + a_1'Z_t + b_0A_t$$

$$Y_{t+1} = a_0 + a_1'Z_t + b_0A_t + b_1'A_tS_t$$

- Regression coefficient interpretation (b_0, b_1)
- Choice of Z_t
- Centered and weighted least squares estimation
- Incorporation of person-specific effects?

Estimated Hazard Ratios (95% Confidence Intervals) for 20% Increments in Total Energy (TE) Consumption and in Activity-related Energy Expenditure (AREE), With and Without Calibration to Correct for Measurement Error, for Various Cardiovascular Disease Categories in the Women's Health Initiative Observational Study (OS) from Baseline (1994-1998) Through September 30, 2010 (Zheng et al, 2014, AJE)

Disease Category	Uncalibrated				Calibrated			
	Energy		AREE		Energy		AREE	
	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI
Total CHD	1.00	0.98,1.02	0.99	0.97,1.01	1.57	1.19,2.06	0.78	0.65,0.95
Nonfatal MI	1.00	0.98,1.03	0.99	0.97,1.01	1.49	1.13,1.97	0.80	0.67,0.97
Coronary Death	0.97	0.94,1.02	0.97	0.94,1.00	2.22	1.36,3.61	0.63	0.46,0.86
Congestive Heart Failure	1.04	1.01,1.08	0.97	0.95,1.00	3.51	2.12,5.82	0.57	0.41,0.79
CABG and PCI	1.01	0.99,1.03	1.01	0.99,1.03	1.43	1.19,1.70	0.90	0.79,1.03
Total Stroke	0.97	0.95,1.00	0.99	0.98,1.01	1.36	1.05,1.76	0.83	0.69,0.99
Ischemic Stroke	0.98	0.96,1.01	0.99	0.97,1.01	1.55	1.14,2.10	0.78	0.64,0.94
Hemorrhagic Stroke	0.94	0.88,0.99	1.03	0.99,1.08	0.47	0.21,1.07	1.37	0.85,2.20
Total CVD: CHD and Stroke	0.99	0.97,1.00	0.99	0.98,1.00	1.49	1.18,1.88	0.80	0.69,0.92
Total CVD including CABG and PCI	1.00	0.99,1.01	1.00	0.99,1.01	1.49	1.23,1.81	0.83	0.73,0.93

Estimated Hazard Ratio (95% Confidence Interval) for 20% Increments in Total Energy (TE) Consumption and in Activity-related Energy Expenditure (AREE), With and Without Calibration to Correct for Measurement Error, for Various Cancer Categories, in the Women's Health Initiative Observational Study (OS) from baseline (1994-1998) Through September 30, 2010 (Zheng et al, 2014,AJE).

Cancer Category	Uncalibrated				Calibrated			
	Energy		AREE		Energy		AREE	
	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI
Total Invasive Cancer	1.01	1.00,1.02	0.99	0.99,1.00	1.43	1.17,1.73	0.84	0.73,0.96
Obesity-related Cancer	1.02	1.00,1.03	1.00	0.99,1.01	1.71	1.33,2.21	0.79	0.65,0.94
Breast Cancer	1.01	0.99,1.02	1.00	0.99,1.01	1.47	1.18,1.84	0.82	0.71,0.96
Colon Cancer	1.00	0.96,1.03	1.00	0.97,1.03	1.86	1.18,2.93	0.83	0.66,1.04
Rectum Cancer	1.01	0.92,1.10	0.99	0.93,1.05	2.75	1.10,6.83	0.51	0.27,0.99
Ovary Cancer	1.00	0.95,1.05	1.01	0.98,1.05	0.85	0.43,1.68	1.12	0.73,1.71
Endometrial Cancer	1.08	1.04,1.12	1.01	0.98,1.05	2.72	1.44,5.13	0.77	0.49,1.21
Bladder Cancer	1.03	0.97,1.10	0.96	0.92,1.00	1.80	0.88,3.69	0.68	0.42,1.09
Kidney Cancer	1.05	0.98,1.12	1.02	0.96,1.07	2.94	1.37,6.28	0.62	0.35,1.12
Pancreas Cancer	0.95	0.89,1.01	0.97	0.92,1.01	2.06	0.98,4.33	0.61	0.37,1.00
Lung Cancer	0.99	0.96,1.01	0.97	0.95,1.00	1.14	0.74,1.76	0.79	0.60,1.03
Lymphoma	1.08	1.03,1.13	1.00	0.96,1.03	0.99	0.48,2.07	1.16	0.69,1.94
Leukemia	1.01	0.95,1.07	0.98	0.93,1.02	1.48	0.70,3.12	0.74	0.47,1.18

Estimated Hazard Ratio (95% Confidence Interval) for 20% Increments in Total Energy (TE) Consumption and in Activity-related Energy Expenditure (AREE), With and Without Calibration to Correct for Measurement Error, for Diabetes Incidence, in the Women’s Health Initiative Observational Study (OS), from Baseline (1994-1998) Through September 17, 2012 (Zheng et al, 2014,AJE).

Outcome Category	Uncalibrated				Calibrated			
	Energy		AREE		Energy		AREE	
	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI
Diabetes Mellitus	1.06	1.04,1.07	1.01	1.00,1.02	4.17	2.68,6.49	0.60	0.44,0.83