# A Sampling of IMPACT Research:
## Methods for Analysis with Dropout and Identifying Optimal Treatment Regimes

Marie Davidian
Department of Statistics
North Carolina State University


http://www.stat.ncsu.edu/∼davidian

# Outline

- Overview: Projects 2 and 5
- Methods for primary and longitudinal analyses in the presence of dropout
- Identifying optimal treatment regimes from a restricted, feasible set
- Computational Resource and Dissemination Core

# Overview

**IMPACT:**

- *P01 Program Project* grant from NCI
- Five *research projects*
- Three *cores*

**Focus here:** Research being carried out in *two* of the projects

- *Project 2*: Methods for Missing and Auxiliary Covariates in Clinical Trials
- *Project 5*: Methods for Discovery and Analysis of Dynamic Treatment Regimes
- *Of necessity*: Simplest cases

# Project 2

**Specific aims:**

1. Improving efficiency of inferences in randomized clinical trials using auxiliary covariates
2. *Methods for primary and longitudinal analyses in the presence of drop-out*
3. Diagnostic measures for longitudinal and joint models in the presence of missing data
4. Inference for sensitivity analyses of missing data

# Doubly robust methods in the presence of dropout

**Motivation:**

- Subject drop-out is *commonplace* in clinical trials
- Particularly problematic in studies of *longitudinal markers*, e.g., QOL measures, biomarkers
- *Monotone* pattern of missingness

**Missing at random (MAR)**: Probability of drop-out depends only on information *observed prior to drop-out*

- *Likelihood methods*: Do not require specification of drop-out mechanism but do require *correct* full data model
- *Inverse weighted methods*: Do not require full data model but do require *correct* drop-out model
- *Doubly robust methods*: Require both, but *only one* need be correct

# Doubly robust methods in the presence of dropout

**Doubly robust methods:**

- *Obvious* appeal
- But "*usual*" doubly robust methods can exhibit *disastrous performance* under "*slight*" model misspecification (Kang and Schafer, 2007)

**Goal:** Can *alternative* doubly robust methods be developed that do not suffer this shortcoming?

# The simplest setting

**Clinical trial:**

- *Outcome* $Y$, interested in $\mu = E(Y)$
- *Full data*: $(Y_i, X_i)$, $i = 1, \ldots, n$, iid, $X_i$ = *baseline covariates* for subject $i$
- But $Y_i$ is *missing* for some $i$ (e.g., due to *drop-out*)
- *Observed data*: $(R_i, R_i Y_i, X_i)$, $i = 1, \ldots, n$, iid, $R_i = I(Y_i \text{ observed})$

**MAR assumption:** $R_i \perp\!\!\!\perp Y_i \mid X_i$, implies

$$\mu = E(Y) = E\{E(Y|X)\} = E\{E(Y|R = 1, X)\} \tag{1}$$

# Estimators for $\mu$

**Outcome regression estimator:** MAR (1) suggests *positing* a model $m(X, \beta)$ for $E(Y|X)$

$$\widehat{\mu}^{OR} = n^{-1} \sum_{i=1}^{n} m(X_i, \widehat{\beta}) \quad \text{for some} \quad \widehat{\beta}$$

- By MAR (1), can use *complete cases* with $R_i = 1$; e.g. *least squares*

$$\sum_{i=1}^{n} R_i \{Y_i - m(X_i, \beta)\} m_\beta(X_i, \beta) = 0, \quad m_\beta(X, \beta) = \frac{\partial m(X_i, \beta)}{\partial \beta}$$

- $\widehat{\mu}^{OR}$ *consistent* for $\mu$ if $m(X, \beta)$ is *correct*

# Estimators for $\mu$

**Inverse propensity score weighted estimator:** *Propensity score* $P(R = 1|X)$

- If $\pi(X)$ is the *true* propensity score, by MAR

$$n^{-1} \sum_{i=1}^{n} \frac{R_i Y_i}{\pi(X_i)} \xrightarrow{p} \mu$$

- *Posit* a model $\pi(X, \gamma)$, estimate $\gamma$ by ML on $(R_i, X_i)$, $i = 1, \ldots, n$

$$\widehat{\mu}^{IPW} = n^{-1} \sum \frac{R_i Y_i}{\pi(X_i, \widehat{\gamma})}$$

- $\widehat{\mu}^{IPW}$ *consistent* for $\mu$ if $\pi(X, \gamma)$ is *correct*

# Semiparametric theory

**Robins et al. (1994):** If the *propensity model* is *correct*, with *no additional assumptions* on the distribution of the data

- All *consistent and asymptotically normal* estimators are asymptotically equivalent to estimators of the form

$$n^{-1}\sum\left\{\frac{R_iY_i}{\pi(X_i,\widehat{\gamma})}+\frac{R_i-\pi(X_i,\widehat{\gamma})}{\pi(X_i,\widehat{\gamma})}h(X_i)\right\} \quad \text{for some function } h(X)$$

- *Optimal* $h(X)$ leading to *smallest variance* (*asymptotically*) is

$$h(X)=-E(Y|X)$$

- Suggests *modeling* $E(Y|X)$ by $m(X,\beta)$, estimating $\beta$, and estimating $\mu$ by

$$n^{-1}\sum\left\{\frac{R_iY_i}{\pi(X_i,\widehat{\gamma})}-\frac{R_i-\pi(X_i,\widehat{\gamma})}{\pi(X_i,\widehat{\gamma})}m(X_i,\widehat{\beta})\right\} \tag{2}$$

# New perspective

**Double robustness:** DR

- Such estimators are *consistent* for $\mu$ if *either* model is *correct*
- *Kang and Schafer (2007)*: *Simulation scenario* where the "*usual*" DR estimator of form (2) with $\beta$ *estimated by least squares* is *severely biased and inefficient* when $m(X, \beta)$ and $\pi(X, \gamma)$ are only "*slightly*" misspecified *or* some $\pi(X_i, \widehat{\gamma})$ are *close to 0*
- $\widehat{\mu}^{OR}$ performed *much better*, even under *misspecification* of $m(X, \beta)$

**Key finding:** With DR estimators, the method for estimating $\beta$ *matters*

- The method that is best for estimating $\beta$ *is not best* for estimating $\mu$
- *Instead*: Find an estimator for $\beta$ that *minimizes* the (large sample) *variance* of DR estimators of form (2)...

# New perspective

**Idea:** Assume $\pi(X)$ fixed (no unknown $\gamma$) and consider estimators

$$n^{-1}\sum\left\{\frac{R_iY_i}{\pi(X_i)}-\frac{R_i-\pi(X_i)}{\pi(X_i)}m(X_i,\beta)\right\} \quad \text{indexed by } \beta \tag{3}$$

- If $\pi(X)$ is *correct* but $m(X,\beta)$ *may not be*, all estimators of form (3) are *consistent* with *asymptotic variance*

$$\text{var}(Y)+E\left[\left\{\frac{1-\pi_0(X)}{\pi_0(X)}\right\}\{Y-m(X,\beta)\}^2\right] \tag{4}$$

- *Minimize* (4) in $\beta \Longrightarrow \beta^{opt}$ satisfies

$$E\left[\left\{\frac{1-\pi_0(X)}{\pi_0(X)}\right\}\{Y-m(X,\beta^{opt})\}m_\beta(X,\beta^{opt})\right]=0 \tag{5}$$

- Find an estimator $\widehat{\beta} \xrightarrow{p} \beta^{opt}$ under these conditions *and* $\widehat{\beta} \xrightarrow{p}$ *true* $\beta_0$ if $m(X,\beta)$ is *correct* even if $\pi(X)$ is *not*

# New perspective

**Result:** Instead of estimating $\beta$ by *least squares* solving

$$\sum_{i=1}^{n} R_i\{Y_i - m(X_i, \beta)\}m_\beta(X_i, \beta) = 0,$$

estimate $\beta$ by a form of *weighted least squares* solving

$$\sum_{i=1}^{n} R_i \left\{ \frac{1 - \pi(X_i)}{\pi^2(X_i)} \right\} \{Y_i - m(X_i, \beta)\}m_\beta(X_i, \beta) = 0 \qquad (6)$$

- Estimating equation (6) has *expectation* (5) when $\pi(X)$ is *correct*
- The resulting $\widehat{\beta}$ *satisfies* the required conditions
- Can be *generalized* to case of $\pi(X, \gamma)$ with $\widehat{\gamma}$ (ML)
- All this extends to *more general* $\mu$ (e.g., treatment effect)

# New perspective

**Details:** My website and

Cao, W., Tsiatis, A.A. and Davidian, M. (2009). Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data. *Biometrika* **96**, 723–734.

- The DR estimator using this $\widehat{\beta}$ *greatly improved* on the "*usual*" DR estimator and exhibited *superior performance* (to $\widehat{\mu}^{OR}$) in the Kang and Schafer and other scenarios

# Longitudinal study

**Extension:** Longitudinal study with *drop-out*

- *Ideally*: Collect data $L_j$ at time $t_j$, $j = 1, \ldots, M+1$
- *Full data*: $\overline{L} = \overline{L}_{M+1} = (L_1, \ldots, L_{M+1})$
- *Dropout*: If subject is *last seen* at time $t_j$, *dropout indicator* $D = j$, observe only $\overline{L}_j = (L_1, \ldots, L_j)$
- *Observed data*: iid $(D_i, \overline{L}_{D_i})$, $i = 1, \ldots, n$
- *Interest*: Parameter $\mu$ in a *semiparametric model* for the full data
- *Full data estimator for* $\mu$: Solve

$$\sum_{i=1}^{n} \varphi(\overline{L}_i, \mu) = 0, \quad E\{\varphi(\overline{L}, \mu)\} = 0$$

- *MAR*: $\text{pr}(D = j|\overline{L})$ depends only on $\overline{L}_j$, $j = 1, \ldots, M+1$
- *Drop-out model*: $\text{pr}(D = j|\overline{L}) = \pi(j, \overline{L}_j)$, $\pi(M+1, \overline{L}) = \pi(\overline{L})$

# Longitudinal study

**If drop-out model correct:** All *consistent and asymptotically normal* estimators for $\mu$ solve

$$\sum_{i=1}^{n}\left\{\frac{I(D_i=M+1)\varphi(\overline{L}_i,\mu)}{\pi(\overline{L}_i)}+\sum_{j=1}^{M}\frac{dM_{ji}(\overline{L}_{ji})}{K_{ji}(\overline{L}_{ji})}\mathcal{L}_j(\overline{L}_{ji})\right\}=0$$

- $dM_{ji}(\overline{L}_{ji})$, $K_{ji}(\overline{L}_{ji})$ are functions of $\pi(j,\overline{L}_j)$
- These estimators are *DR*
- *Optimal* $\mathcal{L}_j(\overline{L}_j)=E\{\varphi(\overline{L},\mu)|\overline{L}_j\}$; *model* by $\mathcal{L}_j(\overline{L}_j,\beta)$, $j=1,\ldots,M$

**Result:** Can derive *optimal* estimator for $\beta$ by analogy to the previous

Tsiatis, A.A., Davidian, M. and Cao, W. (2011). Improved doubly robust estimation when the data are monotonely coarsened, with application to longitudinal studies with dropout. *Biometrics* **67**, 536–545.

# Project 5

**Specific aims:**

1. Learning methods for optimal dynamic treatment regimes
2. *Identifying optimal dynamic treatment regimes from a restricted, feasible set*
3. Inferential methods for dynamic treatment regimes
4. Design of sequentially randomized trials for dynamic treatment regimes

# Optimal treatment regimes from a feasible set

**Motivation:** *Individualized* (*personalized*) treatment

- *Premise*: Different subgroups of patients may respond *differently* to treatments
- Treatment decisions *tailored* to individual patients based on their *characteristics*, *disease status*, *medical history*, etc
- *Ideally*: Use *all* relevant information in *decision rules*
- *Realistically*: Use a key subset of information *feasibly* collected in *clinical practice*, simple-to-implement, *interpretable* decision rules

**Goal:** Methods for estimating such *feasible dynamic treatment regimes* from data from *clinical trials* or *observational databases*

# The simplest setting

**A single decision:** *Two* treatment options

- *Observed data*: $(Y_i, X_i, A_i)$, $i = 1, \ldots, n$, iid
- $Y_i$ outcome, $X_i$ baseline covariates, $A_i = 0, 1$

**Treatment regime:** A function $g : X \to \{0, 1\}$

- *Simple example*: $g(X) = I(X \leq 50)$
- $g \in \mathcal{G}$, the class of *all* such regimes
- *Optimal regime*: If followed by *all patients* in the population, would lead to *best average outcome* among all regimes in $\mathcal{G}$

# Potential outcomes

**Formalize:** $Y^*(1) =$ outcome if patient were to receive 1; similarly $Y^*(0)$

- *Thus*, $E\{Y^*(1)\}$ is the *average outcome* if *all patients* in the population received 1; similarly $E\{Y^*(0)\}$
- *Assume* we observe $Y = Y^*(1)A + Y^*(0)(1-A)$
- *Assume* $Y^*(0), Y^*(1) \perp\!\!\!\perp A|X$ (*no unmeasured confounders*); automatic in a *randomized trial*
- $\Longrightarrow E\{Y^*(1)\} = E\{\, E(Y|A=1,X)\,\}$; similarly $E\{Y^*(0)\}$
- For any $g \in \mathcal{G}$, define

$$Y^*(g) = Y^*(1)g(X) + Y^*(0)\{1-g(X)\} \qquad (1)$$

- *Optimal regime*: Leads to *largest* $E\{Y^*(g)\}$ among all $g \in \mathcal{G}$; i.e.,

$$g^{opt}(X) = \arg\max_{g\in\mathcal{G}} E\{Y^*(g)\}$$

# Optimal regime

(1): $E\{Y^*(g)\} = E\Big[\, E(Y|A=1,X)g(X) + E(Y|A=0,X)\{1-g(X)\}\,\Big]$

$$g^{opt}(X) = I\{E(Y|A=1,X) - E(Y|A=0,X) \geq 0\}$$

- *Thus*: If $E(Y|A,X)$ is *known* can find $g^{opt}$
- *Posit a model* $\mu(A,X,\beta)$ for $E(Y|A,X)$ and *estimate* $\beta$ based on observed data $\Longrightarrow \widehat{\beta}$
- *Estimate* $g^{opt}$ by $\quad \widehat{g}^{opt}(X) = I\{\mu(1,X,\widehat{\beta}) - \mu(0,X,\widehat{\beta}) \geq 0\}$
- "*Regression estimator*"
- *But*: $\mu(A,X,\beta)$ may be *misspecified*, so $\widehat{g}^{opt}$ could be far from $g^{opt}$

**Alternative perspective:** $\mu(A,X,\beta)$ defines a *class* of regimes, *indexed by* $\beta$, that *may or may not* contain $g^{opt}$

- But may be *feasible* and *interpretable*

# Optimal restricted regime

**For example:** Suppose *in truth*

$$E(Y|A,X) = \exp\{1 + X_1 + 2X_2 + 3X_1X_2 + A(1 - 2X_1 + X_2)\}$$

$$\Longrightarrow g^{opt}(X) = I(X_2 \geq 2X_1 - 1)$$

- *Posit* $\mu(A,X,\beta) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + A(\beta_3 + \beta_4 X_1 + \beta_5 X_2)$
- Defines *class* $\mathcal{G}_\eta$ with elements

$$I(X_2 \geq \eta_1 X_1 + \eta_0) \cup I(X_2 \leq \eta_1 X_1 + \eta_0), \quad \eta_0 = -\beta_3/\beta_5,\ \eta_1 = -\beta_4/\beta_5$$

**Thus, in general:** Consider *class* $\mathcal{G}_\eta = \{g(X,\eta)\}$ *indexed* by $\eta$

- Write $g_\eta(X) = g(X,\eta)$
- *Optimal restricted regime* $g_\eta^{opt}(X) = g(X,\eta^{opt})$,

$$\eta^{opt} = \arg\max_\eta E\{Y^*(g_\eta)\}$$

# Optimal restricted regime

**Approach:** *Estimate* $\eta^{opt}$ by maximizing a "*good*" (*DR*) estimator for $E\{Y^*(g_\eta)\}$

- *Missing data* analogy: "*Full data*" are $\{Y^*(g_\eta), X\}$; "*observed data*" are $(C_\eta, C_\eta Y, X)$, where

$$C_\eta = Ag(X,\eta) + (1-A)\{1-g(X,\eta)\}$$

- $\pi(X) = \text{pr}(A=1|X)$; *known* in a randomized trial; otherwise *model* and *estimate* $\pi(X,\widehat{\gamma})$
- $\pi_c(X) = \text{pr}(C_\eta = 1|X) = \pi(X)g(X,\eta) + \{1-\pi(X)\}\{1-g(X,\eta)\}$

# Optimal restricted regime

**Estimator for $E\{Y^*(g_\eta)\}$:**

$$n^{-1}\sum_{i=1}^{n}\left\{\frac{C_{\eta,i}Y_i}{\pi_c(X_i,\widehat{\gamma})}-\frac{C_{\eta,i}-\pi_c(X_i,\widehat{\gamma})}{\pi_c(X_i,\widehat{\gamma})}m(X_i,\widehat{\beta},\eta)\right\} \tag{2}$$

$$m(X,\beta,\eta)=\mu(1,X,\beta)g(X,\eta)+\mu(0,X,\beta)\{1-g(X,\eta)\}$$

- *Consistent* if *either* $\pi(X,\gamma)$ or $\mu(A,X,\beta)$ is *correct*
- *Maximize* (2) in $\eta$ to obtain $\widehat{\eta}^{opt}$

**Current work:**

- Approaches to *maximizing* (2)
- *Simulations*: Almost equals performance of *correct* regression estimator and is *superior* with *misspecified* $\mu(A,X,\beta)$
- *Extension* to *multiple decision points*
- Zhang, Tsiatis, Davidian (2011), *in preparation*

# Computational Resource and Dissemination Core

**Goals:**

- Efficient, robust, reliable code implementing project methodology
- *Creation and dissemination of public-use software* (to be made available on the IMPACT website)
- E.g., R packages, SAS macros, specialized implementations (e.g., FORTRAN, c)

**Programmers:** Scientific programmers at UNC-CH and NCSU dedicated to these activities