

A Statistician's Adventures in Public Health Research: Balance and Design

Thomas E. Love, Ph. D.

thomaslove@case.edu

Center for Health Care Research & Policy

CASE / MetroHealth Medical Center

March 8, 2006

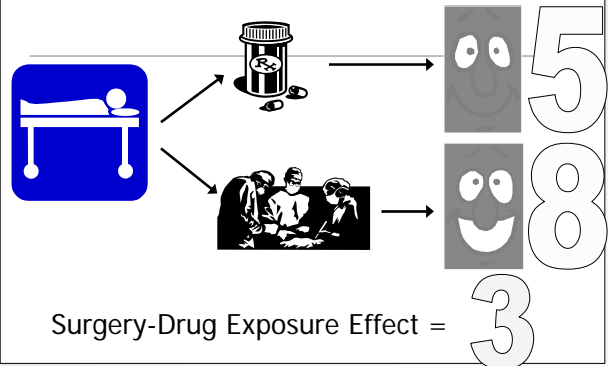
Center for Health Care Research & Policy www.chrp.org

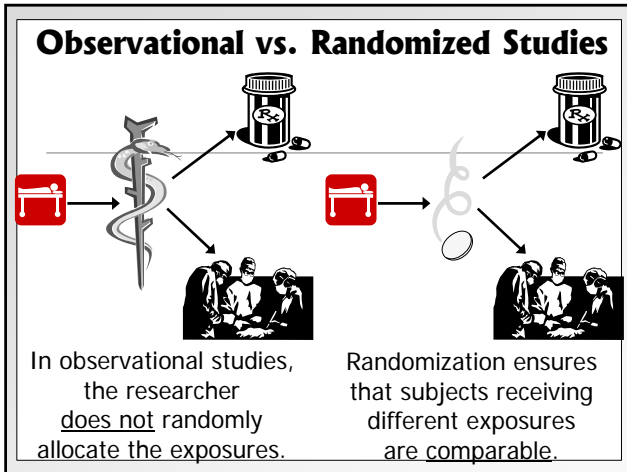


This Afternoon's Examples

1. SUPPORT / Right Heart Catheterization Observational Study
 - How well do propensity score adjustments help to balance "unmeasured" covariates?
2. DIG-IT Cluster Randomized Trial
 - How can cluster-randomized trial design best incorporate detailed information from an electronic medical record?

Looking for Causal Exposure Effects





No Randomization Means ...?

- Goal: Compare groups who looked similar before exposure.
- Can't control exposure assignment, so we can't randomize. Fair comparisons?
 - Analytical adjustments to account for baseline (covariate) differences.
 - Study is biased if exposure groups differ in ways that matter for study outcomes.

Matched Sets / Strata to Adjust for Overt (Selection) Bias

- Observe P covariates, collected in X
 - Even if each is binary, 2^P possible values of X – likely that we'll have unique X values.
- Goal: compare exposed and control groups with similar distributions of X, even if matched subjects have differing X values
- Key tool for doing this: propensity score

Selection Bias and Observational Studies

- Bias: exposure groups differ in ways that matter for the outcomes under study.
 - Overt Bias (seen in data – propensity scores)
 - Hidden Bias (required data not collected)

Propensity Score (PS) = $\Pr(\text{exposure} | \text{covariates } X)$

- Propensity Model: $X \rightarrow$ summary score in $[0,1]$.
- Propensity Score Adjustments account for overt selection bias, but not (directly) for the biasing effects of unmeasured or omitted variables.

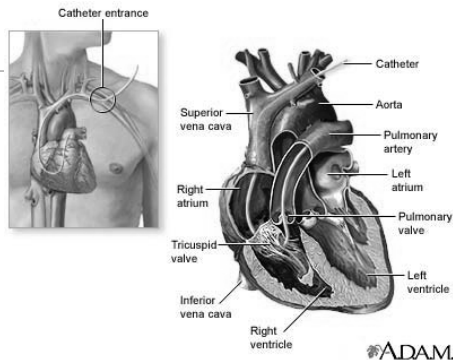
Causal Analysis via PS Match

Received Exposure	.62	.74	.58	.81		
	A	B	C	D		
Add propensity scores based on baseline characteristics	1	2	3	4	5	6
No Exposure	.74	.61	.36	.81	.59	.43

Selection Bias Adjustment and Propensity Score Matching

- Form pairs of subjects matched on PS.
- Use standard analyses for randomized experiments on propensity-matched pairs.
 - Subjects with similar propensity scores analyzed as if assigned randomly.
 - PS matching balances observed covariates
- What about hidden bias?
 - Accounting for “unmeasured” covariates?

Right Heart (Swan-Ganz) Catheterization



<http://www.nlm.nih.gov/medlineplus/ency/imagepages/18087.htm>

Results of Previous (Observational) Studies of RHC

- For RHC patients (as compared to non-RHC patients), relative risk of death:
 - higher in the elderly
 - higher in patients with acute MI
- Patients with higher than expected use of RHC had higher mortality.
- Problem: Selection Bias. Physicians decide who gets RHC.

Why Not Do A Randomized Controlled Trial of RHC?

- Why does someone get a RHC?
 - RHC directly measures cardiac function
 - Some MDs believe RHC is necessary to guide therapy for some critically ill patients
 - Does the RHC do more harm than good?
- Can't do a randomized controlled trial
 - Procedure is very popular
 - Some MDs could not ethically participate

The SUPPORT / RHC Study

Connors AF et al. (1996) *JAMA* 276, 889-897

- Goal: Examine the association between use of RHC during the first 24 hours in the ICU and outcomes.
- Outcomes: survival, length of stay, intensity of care, cost of care
- Sample: 5735 critically ill adult ICU patients in nine disease categories

SUPPORT Disease Categories

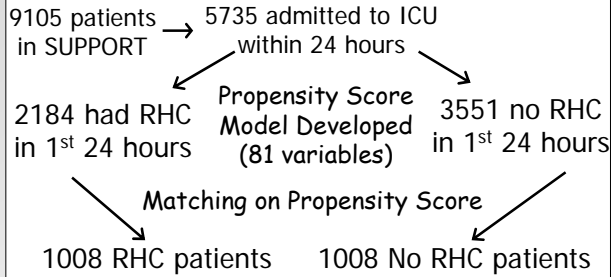
- Acute respiratory failure (ARF)
- Chronic obstructive pulmonary disease (COPD)
- Congestive heart failure (CHF)
- Cirrhosis
- Non-traumatic Coma
- Colon cancer metastatic to the liver
- Non-small cell cancer of the lung (stage III or IV)
- Multi-organ system failure (MOSF) with malignancy or sepsis

Treatment Selection Bias in RHC

- Treatment selection is confounded with patient factors that are related to outcomes
 - Low BP patients are more likely to get RHC
 - Low BP patients are also more likely to die
- Need to identify and measure the variables that independently affect the treatment decision

The SUPPORT / RHC Study

Connors AF et al. (1996) *JAMA* 276, 889-897



A Few of the Covariates in the RHC Propensity Model

- Age, Sex, Race
- Education, Income
- Insurance type
- Primary & secondary disease category
- 12 categories of admission diagnosis
- ADL & DASI 2 weeks before admission
- DNR status on day 1
- Cancer (none, local, metastasized)
- 2-mo survival model
- Weight, temp, BP, heart rate, resp rate
- 13 categories of comorbid illness
- Body chemistry (pH, WBC, PaCO₂, etc.)

Selection Bias in these Continuous RHC Covariates?

($p \leq .001$)

Variable	No RHC	RHC
Sample size	3551	2184
APACHE III Score	51 (38, 62)	61 (47, 74)
2-mo survival model	0.61 (0.49, 0.76)	0.56 (0.45, 0.72)
# comorbid illnesses	1.7 (1, 3)	1.6 (1,2)
Heart rate, bpm	112 (76, 140)	119 (105, 145)
Blood pressure	85 (53, 119)	68 (47, 73)
Creatinine	168 (80, 177)	221 (106, 265)

Mean (25th percentile, 75th percentile)

Categorical RHC Covariates?

($p \leq .001$ except Race)

Variable	No RHC	RHC
Male	1914 (54)	1218 (59)
Disease: MOSF	1245 (35)	1235 (57)
Age >80	500 (14)	167 (8)
DNR status, day 1	710 (20)	296 (14)
Any cancer?	899 (25)	457 (21)
Race non-white	798 (22)	477 (22)

- Mean PS for RHC patients was .577, mean PS for no RHC patients was .253

Characteristics of RHC Patients

- RHC patients were more likely to ...
 - Be male, have private insurance, enter the study with ARF, MOSF or CHF
- RHC patients were less likely to ...
 - Be over 80 years old, have cancer, have a DNR order in the first 24 hours of hospitalization
- RHC patients had significantly...
 - Fewer comorbid conditions, more abnormal results of vital signs, albumin, creatinine, etc.
 - Lower model probability of 2-month survival

Propensity Score Weighting Approach

- An exposed subject's weight is the inverse of its propensity score.
$$w_i = \frac{1}{PS_i}$$
- An unexposed subject's weight is the inverse of 1 minus its PS.
$$w_i = \frac{1}{(1 - PS_i)}$$

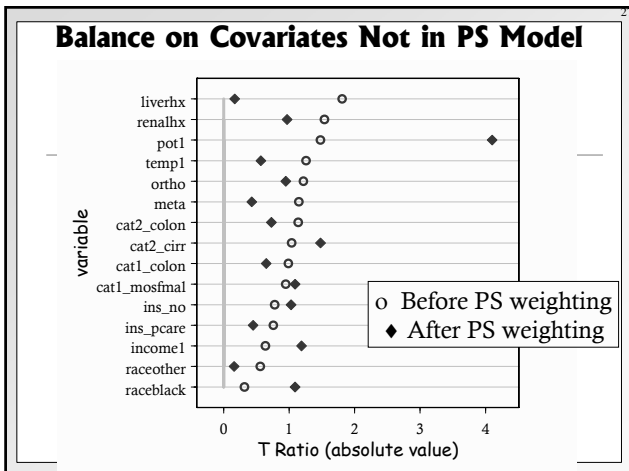
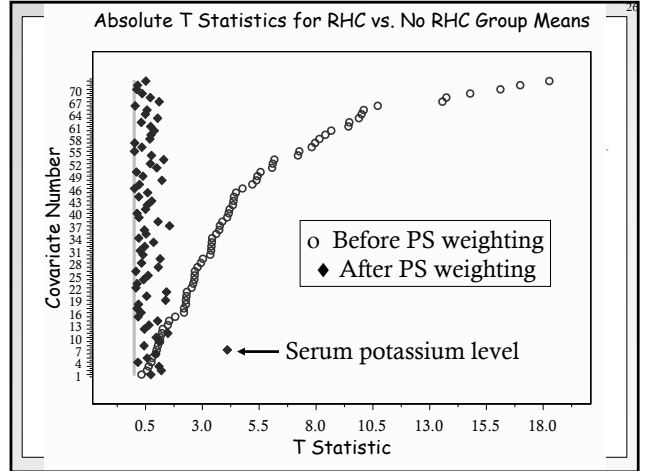
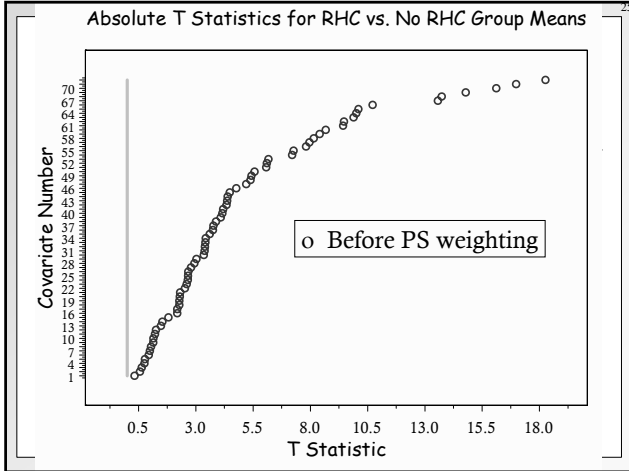
Rubin (2001). Also see Lunceford and Davidian (2004)

Weighting and SUPPORT/RHC (Hirano & Imbens, 2001)

- 5735 patients in the study: 2184 treated (RHC) and 3551 controls (no RHC).
- Outcome: indicator of 30 day survival
- A subset of 72 covariates were considered in this analysis.
- Developed PS using 57 of the 72 covariates (all with t statistics > 2.0).

Covariate Balance Improved by Propensity Score Weighting?

- Build PS model with 57 covariates
 - Reweight each RHC observation by the inverse of estimated PS, and each non-RHC observation by the inverse of 1 minus the estimated propensity score.
- Weighting brings most covariates' means closer together, but not ALL...



Effectiveness of RHC Propensity Score Weighting

- The weighting is based on a propensity model including 57 of the 72 covariates (serum potassium not included).
 - Most means are much closer, although 6 become less balanced after weighting.
 - None of these 6 were in the 57-variable PS model. Err on the side of inclusion...
- PS Weighting appears to balance control and treatment groups well.

Propensity Score Matching Describing Covariate Balance

Before Match (n = 5735) After Match (n = 2016)

Variable	RHC	noRHC	p	RHC	noRHC	p
% Admit-ICU	81.9	57.8	<.001	78.8	77.8	.63
Mean BP	68.20	84.87	<.001	71.18	72.53	.38
Serum Na	136.33	137.04	.001	136.54	136.49	.85

- Balance of these covariates appears much improved after match. All 81 improve.

Standardized Differences To Assess Covariate Balance

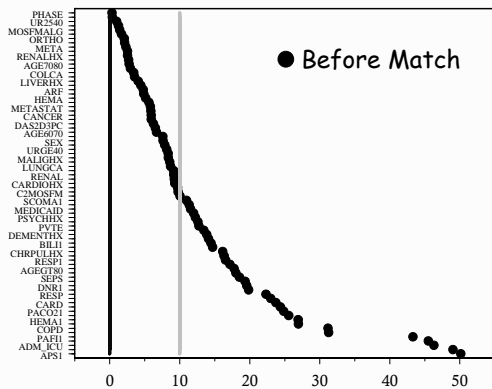
Before Match After Match

Variable	RHC	noRHC	StzD	RHC	noRHC	StzD
AdmitICU	81.9	57.8	46.3	78.8	77.8	2.0
Mean BP	68.2	84.9	-45.5	71.2	72.5	-3.9
Serum Na	136.3	137.0	-9.2	136.54	136.49	0.8

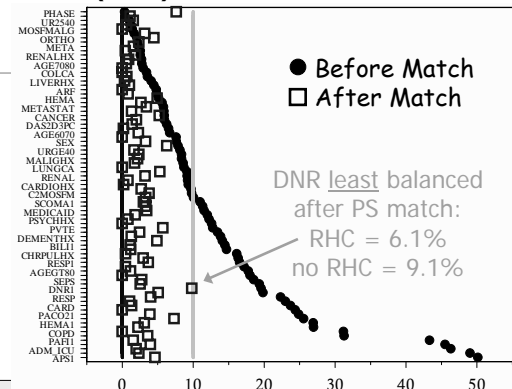
$$\text{StzD} = 100 \left(\bar{x}_{Exp} - \bar{x}_{UnExp} \right) / \sqrt{\frac{s_{Exp}^2 + s_{UnExp}^2}{2}}$$

- |StzD| > 10% indicates poor balance.

Absolute Standardized Differences (in %) for 81 RHC Covariates



Absolute Standardized Differences (in %) for 81 RHC Covariates



Propensity Matching in RHC Balances Observed Covariates

- Before match, 38 of 81 (47%) variables in PS model showed serious imbalance (absolute stand'zd diff. exceeds 10%).
- After the match, none of the 81 variables display serious imbalance. Mean bias reduction is 36% across all 81 variables.
- For the 38 “poorly balanced”... Mean bias reduction was 87.7%, and the range of reductions was 49.9% to 100%.

Looking at “Unmeasured” Variables

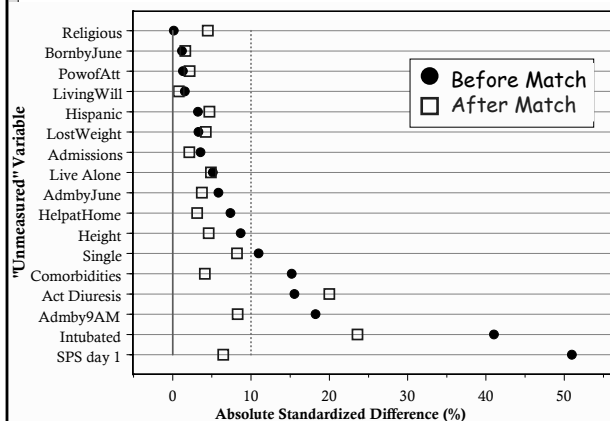
- Identified 17 covariates from SUPPORT that were not in the propensity model:
 - Likely to show differences by exposure group (RHC/no RHC) or an impact on outcomes (e.g., # of comorbidities)
 - Unlikely to show a relation to either exposure or outcome (e.g., birth month)
- 11 of the 17 variables were well balanced ($|\text{stdzd diff.}| < 10\%$) before matching.

17 “Unmeasured” Covariates

✓ Significantly ($\alpha = .05$) correlated with PS model

- | | |
|---|-----------------------------------|
| <input type="checkbox"/> Not religious? | ✓ Help at home? |
| <input type="checkbox"/> Born in Jan – June? | ✓ Patient's height |
| <input type="checkbox"/> Power of attorney? | ✓ Marital status (Single?) |
| <input type="checkbox"/> Understands Living Will? | ✓ # of comorbid illnesses |
| <input type="checkbox"/> Hispanic? | ✓ Active diuresis, day 1 |
| <input type="checkbox"/> Lost Weight past 2 wks? | ✓ Admitted by 9 AM? |
| ✓ # hospital admissions | ✓ Intubated, day 1 |
| ✓ Lives alone? | ✓ Support Physiology Score, day 1 |
| <input type="checkbox"/> Admitted Jan – June? | |

17 “Unmeasured” Covariates



Does Propensity Match Balance “Unmeasured” Covariates?

Support Physiology Score (SPS)

Before Match After Match

	RHC	noRHC	StzD	RHC	noRHC	StzD
SPS	37.41	31.90	51	35.20	35.92	-6

- Bias due to SPS is reduced by $(51-6)/51 = 88\%$.
- SPS significantly correlated with PS ($r = .44$)

Correlation with PS?	# of Covariates	Balance Improved Post-Match?	Median Bias Reduction
Sig. ($\alpha = .05$)	10	9 (90%)	45%
Not Significant	7	2 (29%)	-36%

Impact on a Key Outcome

PS Model	Matched Pairs	30d Mortality Odds Ratio	95% CI for Odds Ratio
81 variables	1008	1.23	(1.04, 1.43)
98 variables	1252	1.21	(1.02, 1.44)

- No detectable impact of “unmeasured” covariates on conclusions after PS match.

What Makes an Unmeasured or Omitted Variable Dangerous?

- All observational studies are potentially affected by hidden bias. Sensitivity analyses are a necessary part of any such study.
- An omitted variable is most likely to affect conclusions about the exposure if it is:
 - closely related to outcome.
 - seriously imbalanced by exposure.
 - uncorrelated with propensity score.

This Afternoon’s Examples

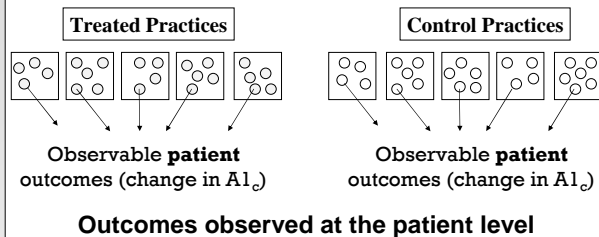
- SUPPORT / Right Heart Catheterization Observational Study
 - How well do propensity score adjustments help to balance “unmeasured” covariates?
- DIG-IT Cluster Randomized Trial
 - How can cluster-randomized trial design best incorporate detailed information from an electronic medical record?

Cluster-Randomized Trials and Electronic Medical Records

- Trials where clusters of patients, rather than individuals, are randomized to interventions.
- CRTs are increasingly important in public health research.
- Electronic medical record (EMR) data very useful in executing well-balanced designs for health intervention trials.

What Is A Cluster-Randomized Trial?

Randomization is at the practice level



The Cluster Randomized Trial

- Unit randomized not the only unit of interest
- Interventions focus on systems, behavior
- Useful when “contamination” is likely to be a problem and/or blinding is not possible
- Members of clusters are likely more similar to each other than to members of other clusters
 - This affects study design, power and “effective” sample size, analytic strategy

Diabetes Improvement Group – Intervention Trial [DIG-IT]

A Cluster Randomized Trial of Electronic Decision Support on Improvement of Diabetes Care and Outcomes

Key DIG-IT Design Parameters

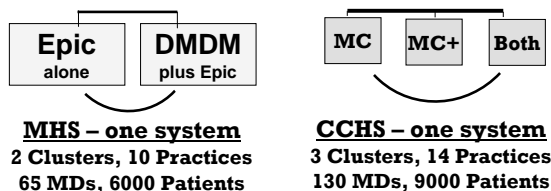
- Primary Interventions
 - different approaches to decision support (DMDM for primary care providers, MyChart and MyChart+ for patients)
- Primary outcome of interest
 - difference in changes across study groups in hemoglobin A1c values.

DIG-IT focuses on the Donabedian Triad

- Structure
 - Health systems with Epic ambulatory electronic medical records
- Process
 - Improving care through decision support
 - Disease Management for diabetes mellitus DM²
 - Patient Empowerment through MyChart = MC
 - Alone or Enhanced for glycemic control = MC+
- Outcomes
 - Measures of Diabetes Control (broad set)
 - Utilization and Costs

What is DIG-IT?

CRT of **physician** (diabetes mellitus disease management) and/or **patient** (MyChart, MyChart-plus) decision support underway in MetroHealth & Cleveland Clinic Health Systems



Pre-Randomization Balancing in the Pre-EMR Era

- Pre-EMR Era:
 - Relatively meager amounts of information known about cluster members before the study:
 - e.g.; geography; average age (maybe); distribution of race and sex, perhaps 1-2 other prognostic factors at best
 - Logistical concerns drive the design.

Importance of Balancing in CRTs

- Investigators conducting a CRT want minimal cross-cluster differences on important predictors of response to the intervention; as well as on other covariates associated with outcomes.
- DIG-IT's 10 MHS practices draw from varied populations; thus balancing of clusters was critical for developing fair and statistically powerful comparisons.

Pre-Randomization Balancing in the Era of Electronic Medical Records

- Match or stratify practice sites on substantial historical information, then randomly allocate the exposures to practice sites.
 - predictors of response to intervention
 - historical measures of outcome(s)
- We can't balance all the variables, but we can investigate all the "clusterings."

DIG-IT Procedure to Balance MetroHealth System Clusters

- For all possible clusterings of 10 practices in 2 clusters...
 - Assemble practice-level clinical and demographic data from the EMR
 - Identify clusterings which appear to balance an array of baseline characteristics, including trends in selected diabetes-related parameters.
- Investigators came to a (blinded) consensus as to which option provided best balance.
- The intervention was then randomly allocated to one cluster from that clustering.

Which "Table 1" Do You Want?

Group	#204		#457	
	A	B	A	B
# Practices	4	6	5	5
# Patients	1499	2807	2281	2025
# Physicians	24	68	63	29
Mean slope A1c	-.042	-.055	-.047	-.053
% last A1c >= 9	19.6	18.2	19.8	17.4
Mean last A1c	7.61	7.47	7.64	7.39
Mean last SBP	138.3	135.0	135.7	136.7
% on Insulin	24.0	16.3	19.5	18.5

Which "Table 1" Do You Like?

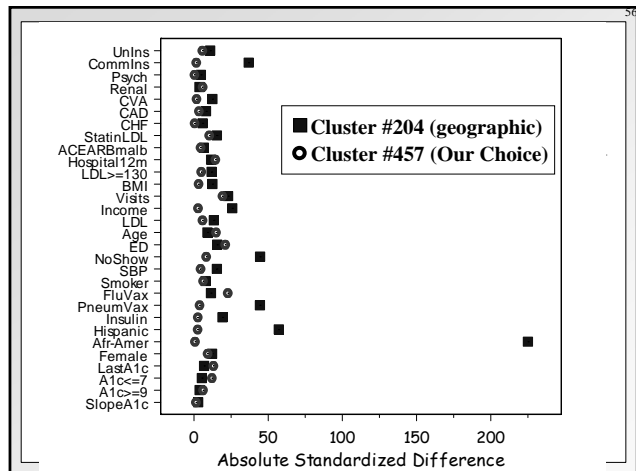
Group	#204		#457	
	A	B	A	B
% Female	68.9	63.0	67.1	62.7
% African-Am.	96.6	23.3	48.7	49.1
% Hispanic	0.4	15.3	9.7	10.5
% Pneumovax	83.4	64.4	71.5	73.2
% Flu vax 12m	30.4	35.8	38.8	28.1
% curr smoker	26.2	22.8	25.2	22.6
% >=1 No Show	65.8	44.2	49.8	53.9
% >=1 ED Visit	23.3	30.3	32.2	22.8

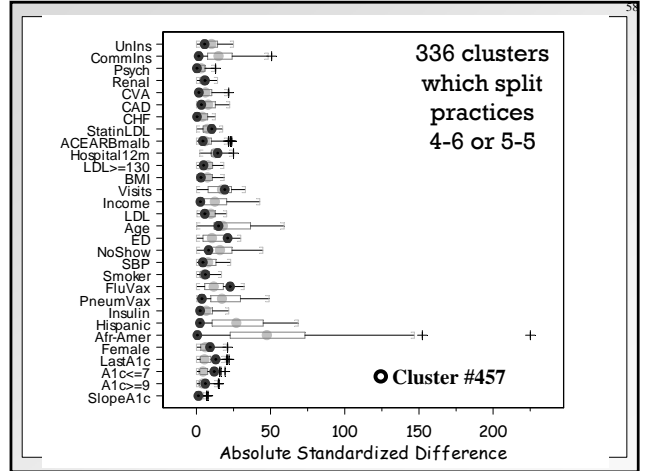
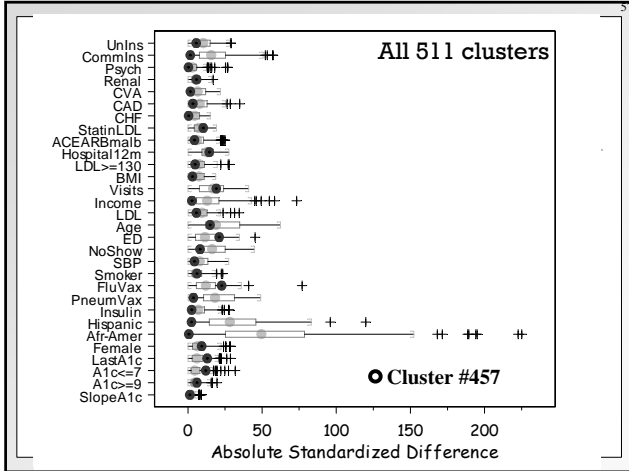
Choose Your "Table 1"

Group	#204		#457	
	StzDiff	P	StzDiff	P
Mean slope A1c	2.9	0.39	1.4	0.67
% last A1c >= 9	3.8	0.25	6.1	0.05
Mean last SBP	15.5	<0.001	4.4	0.15
% on Insulin	19.3	<0.001	2.5	0.41
% Female	12.3	<0.001	9.2	0.003
% African-Am.	225.1	<0.001	-0.7	0.83
% Pneumovax	44.4	<0.001	-3.8	0.27
% curr smoker	7.9	0.01	6.1	0.05

#457: "Not Deliberately Balanced"

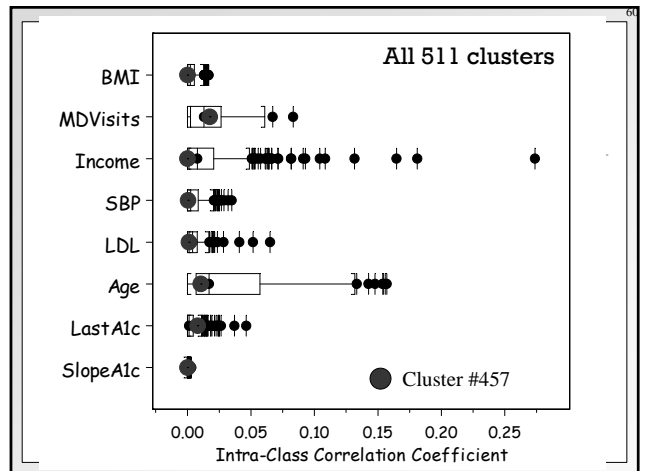
Group	A	B	StzDiff	P
Mean Age	57.6	59.6	-14.9	<0.001
Mean last BMI	34.30	34.57	-3.1	0.30
Mean LDL	115.5	117.7	-5.7	0.07
Mean income	36377	36899	-2.7	0.40
Mean MD visits	4.09	4.64	-19.1	<0.001
% Hosp in 12 m	19.5	14.1	14.4	<0.001
% CHF	6.9	6.8	0.5	0.90
% malb ACE/ARB	36.0	33.9	4.5	0.62
% hiLDL Statin	41.9	36.9	10.3	0.07





Intra-Class Correlation Coefficients for Continuous Variables

	#457	Range [4-6,5-5]
Slope of Hemoglobin A1c	<0.001	<0.001, 0.003
Last Hemoglobin A1c	0.008	<0.001, 0.025
Last Systolic BP	0.001	<0.001, 0.025
Age	0.011	<0.001, 0.148
Last LDL	0.001	<0.001, 0.021
Median (block) 1999 Income	<0.001	<0.001, 0.093
# MD Visits	0.018	<0.001, 0.052
Last BMI	<0.001	<0.001, 0.017



DIG-IT Balancing for MHS

- Our allocation yielded 5 control practice sites and 5 practice sites in the intervention group.
- Excellent Balance achieved for numerous baseline characteristics.
- Balance substantially superior than that of groupings that did not use detailed EMR data.

CRT and EMR: Conclusions

- EMR systems provide new opportunities for state-of-the-art study design.
- Our approach could be effectively used to create study groups for a wide range of community-based therapeutic trials or health care delivery trials.