

CIRA Methodology and Biostatistics Seminar Series

“Estimation of Causal Effects in Observational Studies”

Marilyn Stolar, Ph.D.,
CIRA Methodology & Biostatistics Core
Co-Director for Data Management & Statistical Analysis
Yale University

The randomized controlled experiment is considered the “gold standard” design when researchers attempt to estimate the *causal* effect of a treatment, exposure, or risk factor. However, in HIV/AIDS-related research, it is often the case that subjects cannot be randomized to the comparison groups of interest, thereby introducing bias which threatens the validity of the estimates of causal effect parameters.

In the context of examples, various statistical methods for mitigating potential bias will be examined. We begin with a brief overview of standard methods, stratification and adjustment for confounders. We will then consider matching and propensity scoring, multiprocess modeling, and marginal structural modeling approaches. Strengths and weaknesses of each method will be discussed.



Thursday, February 19th, 2004
2:15 - 3:45 @ CIRA Conference Room
40 Temple Street, Suite 1B

Yale University Center for Interdisciplinary Research on AIDS

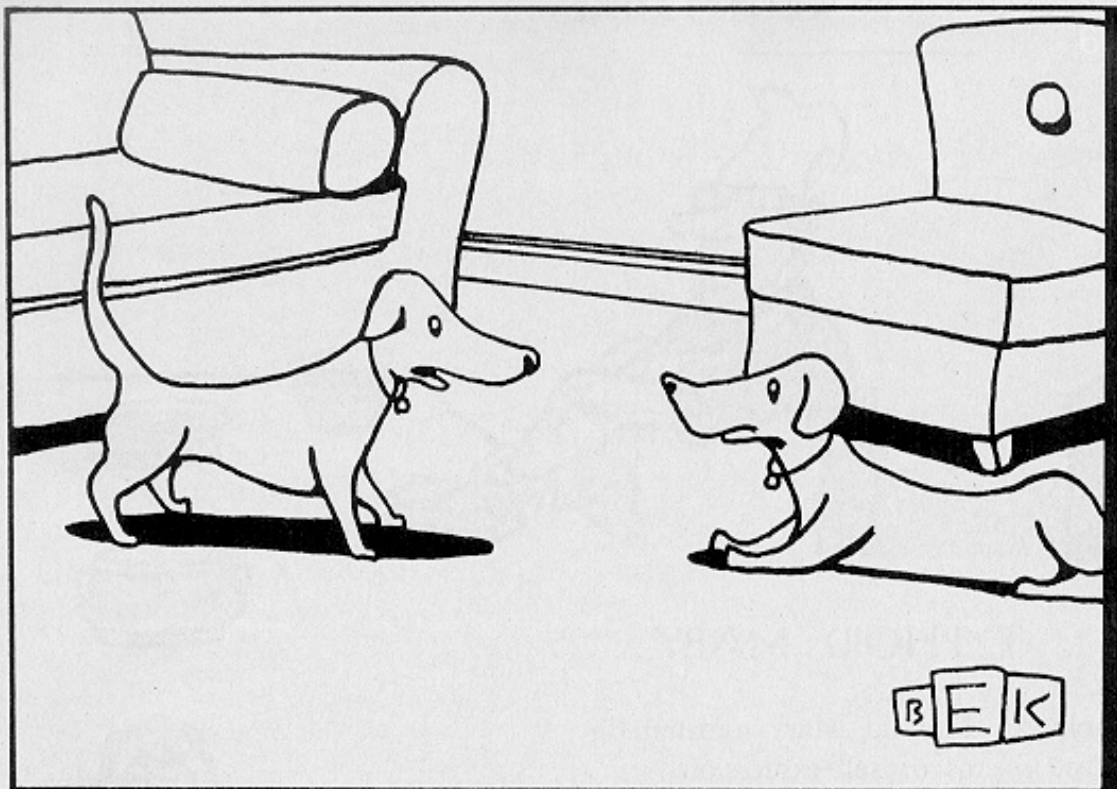
Cause and Effect

In this presentation, we are not considering the substantive scientific burden of proof that must be met to establish that one event is the cause of another. Rather, we grant that relationship, and then discuss statistical issues pertaining to the estimation of effect size under an observational study design.

Measuring Causality

- Ideal—each subject would be observed under each condition; their responses would be differenced and combined to get an average causal effect

--counterfactual world--



"It's the naps you don't take that you regret the most."

Measuring Causality

- Next best—randomization—two nearly identical groups differ only in their assigned condition—difference their averages to get average difference

Observational Study

- Individuals are observed already under their respective conditions
- The two groups may differ in ways other than just the observed condition
- Average effects may be biased due to confounding between covariates and group condition.

Bias Due To Confounding

Unless...

We can simulate randomization or counterfactual world using information from observational study.

Standard Remedies

- Matching
- Stratification on covariates
- Statistical models with covariate adjustment

Propensity Score

- Defined as “...the conditional probability of being in the treatment, exposure or risk group of interest given the observed covariates...”
- Applied to pairs of groups.

Propensity Score

- Main idea: replace the collection of confounding covariates with a single value which is a function of these covariates representing the 'propensity' to receive the treatment.
- This score is then used as if it were the only confounding covariate.

Calculating the Propensity Score

- The propensity score may be found, for example, via logistic or probit regression, or discriminant analysis.
- Dependent Variable = 'treatment' / 'control'
--dichotomous--
- Explanatory Variables = set of potential confounding covariates

Calculating the Propensity Score

- Explanatory Variables= set of potential confounding covariates
 - > nuisance variables that are unassociated with group or with the outcome of interest are not a problem

Calculating the Propensity Score

- Explanatory Variables= set of potential confounding covariates
 - > it is critically important that the outcome variable (or a virtual proxy) not be among the set of potential confounding covariates used to calculate the propensity score

Properties of the Propensity Score

- Two subjects with different covariate profiles can have the same propensity score.
- The propensity score is a 'coarse' function of the covariates that serves as a 'balancing score'.
- Matching, stratification or covariance adjustment using the propensity score balances the observed covariates between the two treatment groups.

Properties of the Propensity Score

- Matching, stratification or covariance adjustment using the propensity score balances the observed covariates between the two treatment groups creating a “quasi-randomized experiment”.
- Propensity score adjustment reduces bias and increases the precision of (power to detect) the estimate of the treatment effect.

Interesting Result

Theoretical results indicate that as long as there is 'sufficient' overlap between the propensity score distributions of the T and C groups, stratification by 5 or 6 subclasses of the propensity score will typically remove 90% or more of the bias in the unadjusted treatment-outcome estimate of effect.

Interesting Result

Simulations have shown that propensity score adjustment is likely to balance 'unobserved' covariates as well, although there is no theoretical basis for quantifying chance as there is with randomization.

Procedure

- Identify potential confounders. Be liberal, imaginative.
- Estimate propensity scores for each subject.
- Match, or separate the subjects' propensity scores into quintiles. Verify that each covariate has been successfully balanced* between the T and C groups.

*equality of means, variances, correlations w/in quintiles

Possible Problems

- It may be the case that some covariates do not overlap sufficiently between T and C groups to allow for adjustment.
- It may be the case that the extreme quintiles have a small number of T or C subjects.

Procedure

- If the covariates are adequately balanced, then perform the matched analysis, the stratified analysis, or model the covariate adjustment.
- Which strategy is ‘best’?
- More than two groups—apply pairwise.

Examples

- Post-term birth effects on neuropsychiatric, social and academic achievement in 5-10 yr old children
- covariates: sex of child, parity, mom's age, mode of delivery, child's age, birthweight, race, SES, complications scales, vaginal bleeding, abnormal labor, ...

Examples

- Migration study— first generation Puerto Ricans living in Hartford/New Haven/Bridgeport vs San Juan residents whose ancestors have never migrated
- Effect of Clozapine treatment for schizophrenia on 4-yr suicide risk
- Biological sex effects on math achievement—confounded by gender

Strengths & Limitations

- Strengths—
 - Versus standard methods, the variable reduction (power) and the potential “quasi-randomization” (good bias reduction)
 - Can use many covariates
 - Can reveal serious confounding that would otherwise go unnoticed
 - Easy to perform with standard software and basic statistical tests and models
 - Some aspects are grounded in theory; can do sensitivity analysis

Strengths & Limitations

- Limitations
 - Versus randomization, only accounts for observed variables
 - Works best in larger samples

Marginal Structural Models

Counterfactual logic – estimate the mean/rate of the outcome if everyone in the sample were treated; estimate the mean/rate of the outcome if everyone in the sample were a ‘control’; the differential is the causal effect of treatment on the outcome.

Marginal Structural Models

- Based on the concept of ‘inverse-probability-of-treatment’ weighting in lieu of stratification by confounders
- Takes a ‘missing data’ approach to the counterfactual assignments and outcomes

Procedure

- Calculate $p = \text{the prob of } T \text{ given covariates}$
- For each treated subject, assign them a weight of $1/p$.
- For each control subject, assign them a weight of $1/(1-p)$.
- Perform inverse-probability-of-treatment-weighted analysis of the treatment-outcome model.

Illustration

Consider one covariate:

- 120 subjects
- 48 male, 72 female
- 1/3 of males in T, 1/2 of females in T

16 (48) X3	36 (72) X2	52 T (120)
32 (48) X3/2	36 (72) X2	68 C (120)
48 male	72 female	120 all

Marginal Structural Model

- Especially useful in longitudinal studies with respect to self-selected missingness after randomization, or if treatment regimen adapts to changes in outcome as study evolves.
- Weights can vary for each observation time

Example

- Causal effect of Zidovudine on the survival rate over time of HIV+ men
- Treatment varies with CD4 count

Strengths and Limitations

- Strengths
 - Theoretically grounded
 - Flexible and powerful
- Limitations
 - Relies on observed covariates
 - Not easy to perform

Multiprocess Models

Generally, models with two or more substantive “outcomes” such that at least one outcome may be on the causal pathway for the other.

Example

- In a particular locale, each mother chooses whether or not to deliver each of her babies in a hospital (as opposed to at home or elsewhere).
- We are interested in whether delivering out of hospital increases infant mortality risk.

- Data are collected on 1060 births by 501 mothers during a pre-specified time period.

- Measurements include:
 - mom_id, baby_id, hospital delivery (yes/no), baby's sex ($\text{♀}=0/\text{♂}=1$), mom's education (3 levels), distance of home from nearest hospital, baby's subsequent survival in years, and whether the baby's full length of life is observed or 'censored'.

Standard Model

$$\ln(\mathit{hazard}_{tij}) = \ln(\mathit{baseline hazard}_t) +$$
$$\beta_1 \mathit{dropout}_j + \beta_2 \mathit{college}_j +$$
$$\beta_3 \mathit{boy}_{ij} + \beta_4 \mathit{hospital}_{ij} + u_j$$

$$u_j \sim N(\mathbf{0}, \sigma_u^2)$$

Outcome = mortality 'hazard'
hospital(yes/no) is self-selected risk factor
not significant!

dropout	0.2919= β_1 (1.4894= t)
college	-1.7943 * (2.3729)
boy	0.1880 (1.1811)
hospital	-0.3817 (1.6923)
Var(u_j)	0.5887 ** (2.6239)

(baseline hazard not shown)

Model Considerations

There is significant evidence that the unobserved mother-specific hazard characteristics, represented by u_j , affect their respective children's mortality risk.

The model assumes that the explanatory variable *hospital* is uncorrelated with u_j .

Model Considerations

But suppose that the mothers who believe themselves to be at greater risk of losing their baby are the ones who decide to reduce their risk by delivering in hospital.

Then *hospital* is correlated with residual mother-specific mortality risk u_j .

Model Considerations

Correlation between explanatory variables and residuals leads to biased estimates of the model parameters.

The correlation is due to the self-selection of the comparison groups in this observational study.

Model Considerations

Thus, modeling the correlation between *hospital* and u_j will mitigate the bias.

This is achieved by modeling jointly for each mother the processes of child survival and the decision to deliver in hospital.

The Multiprocess Model

We model the mother-specific effects for each process as draws from a multivariate normal distribution.

By allowing the two mother-specific propensities to co-vary, the bias that was due to the *hospital* and u_j correlation in the standard model is now controlled.

The Multiprocess Model

$$(i) \ln(\mathit{hazard}_{tij}) = \ln(\mathit{baseline hazard}_t) + \beta_1 \mathit{dropout}_j + \beta_2 \mathit{college}_j + \beta_3 \mathit{boy}_{ij} + \beta_4 \mathit{hospital}_{ij} + u_j$$

$$(ii) \mathit{logit}(\mathit{hospital}_{ij}) = \gamma_0 + \gamma_1 \mathit{dist} + \gamma_2 \mathit{dropout} + \gamma_3 \mathit{college} + v_j$$

$$(u_j, v_j) \sim \mathit{MVN}(\mathbf{0}, \mathbf{0}, \sigma_u^2, \sigma_v^2, \rho_{uv})$$

Model Comparison

Hazard:

	<i>standard</i>	<i>multiprocess</i>
dropout	0.2919 (1.4894)	0.2139 (1.0166)
college	-1.7943 * (2.3729)	-1.7326 * (2.2406)
boy	0.1880 (1.1811)	0.1805 (1.1295)
hospital	-0.3817 (1.6923)	-0.6648 * (2.2560)
Var(uj)	0.5887 ** (2.6239)	0.6328 ** (2.8705)

Hospital:

	<i>standard</i>	<i>multiprocess</i>
Constant	-0.0527 (0.2791)	-0.0627 (0.3237)
distance	-0.0623 * (2.1533)	-0.0618 * (2.0487)
dropout	-1.8023 *** (7.6580)	-1.7889 *** (7.4973)
college	0.9898 ** (2.6101)	1.0027 * (2.5860)
Var(vj)	1.1508 *** (6.7252)	1.1630 *** (6.5734)

Rho

**0.5041
(1.6119)**

Results

- With the multiprocess model, the effect of delivering in hospital significantly reduces the mortality hazard.
- The fact that higher risk deliveries are performed in hospital is a source of bias that is controlled in the multiprocess model.

Results

- We find a positive correlation coefficient between unobservables affecting the hospital decision and the mortality risks ($\rho=.52$).
- This suggests that women with above-average risks of losing a baby also tend to have above-average propensities to deliver in hospital (and vice-versa).

Summary

- In this example, need to have multiple observations within subject (e.g. mom)
- Controls for bias induced by non-random assignment in an observational study when the selection 'propensity' is associated with the outcome 'propensity'
- aML software is designed for Multiprocess Multilevel Modeling

(<http://www.applied-ml.com>)

Conclusion

- The statistical enterprise approaches truth by mitigating uncertainty.
- Imperfect, but practical.

Some References

- General
 - Rosenbaum PR, *Observational Studies*, 2nd ed, ©2002, Springer-Verlag, New York
 - Pearl J, *Causality: Models, Reasoning, and Inference*, ©2000, Cambridge University Press, Cambridge UK
- Propensity Scores
 - Rosenbaum P, Rubin DB, *The central role of the propensity score in observational studies for causal effects*, *Biometrika* 1983, v70: 41-55
 - Rubin DB, *Estimating causal effects from large data sets using propensity scores*, *Annals of Internal Medicine* 1997, v127(8S): 757-763
 - D'Agostino Jr RB, *Tutorial in Biostatistics: propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group*, *Statistics in Medicine* 1998, v17: 2265-2281

Some References

- Marginal Structural Models
 - Robins JM, Hernán MÁ, Brumback B, *Marginal structural models and causal inference in Epidemiology*, *Epidemiology* 2000, v11(5): 550-560
 - Hernán MÁ, Brumback B, Robins JM, *Marginal structural models to estimate the causal effect of Zidovudine on the survival of HIV-positive men*, *Epidemiology* 2000, v11(5): 561-570
 - Cook NR, Cole SR, Hennekens CH, *Use of a marginal structural model to determine the effect of aspirin on cardiovascular mortality in the Physician's Health Study*, *American Journal of Epidemiology* 2002, v155(11): 1045-1053
- Multiprocess Modeling
 - Pindyck RS, Rubinfeld DL, *Econometric Models and Economic Forecasts*, ©1991, McGraw-Hill Inc., New York
 - Lillard LA, Panis CWA, *aML Multilevel Multiprocess Statistical Software, Version 2.0*, EconWare, Los Angeles CA