Causal Inference from Epidemiologic Data

Chapter 7. Studies with multiple partially controlled factors.

1 Many studies can be formulated as "controlling" some factors, but not all.

(1) Why ITT analysis can be invalid for ITT effect?

Consider a study with both non-compliance and missing outcomes.

Note 1. The outcome can be censored by non-random reasons.

Note 2. If compliance status C_i predict both the outcome $Y_i(Z)$ and the censoring $R_i(Z)$, then generally $Y_i(Z)$, $R_i(Z)$.

- ITT effect: a causal effect of assignment on all units (without reference to compliance).
- ITT analysis: an analysis that does not use the data on compliance.

An ITT analysis can be invalid for the ITT effect (e.g., Frangakis and Rubin, 1999).

(2) Estimation of causal effect.

(a) Settings as in Chapter 6 for the outcome.

Additional assumptions.

- *Compound exclusion*. If $D_i(1) = D_i(0)$, then $Y_i(1) = Y_i(0)$ and $R_i(1) = R_i(0)$.
- *Latent ignorability*. $Y_i(Z) \coprod R_i(Z) \mid C_i$ and observed covariates.

(b) Results.

- The ITT estimator is inconsistent for ITT effect.
- The ITT effect is estimable consistently if we use compliance data.
- The effect on compliance is not estimable by standard instrumental variables, but it is estimable by using data on missingness.

(c) Modelling.

- pr$(C_i \mid X_i)$.
- pr$(Y_i(z) \mid X_i, C_i)$.

- pr$(R_i(z) \mid X_i, C_i)$.

Case study. Barnard, Frangakis, Hill and Rubin (2003).

2 Partially controlled studies.

We consider studies with

- a controlled factor Z (e.g., new treatment for HIV vs. standard treatment);
- outcome $Y_i(Z)$;
- intermediate variable measured after Z before Y, $S_i(Z)$ (e.g., CD4 counts).

(1) What are causal effects of interest?

E.g., we would be interested in question: is an effect of treatment on the outcome occurring only when an effect of the treatment on the intermediate variable occurs?

Generally interested in causal effects that are also functions of the intermediate outcomes.

(2) Standard definitions: a "net-treatment effect" (Rosenbaum (1984) is defined as a comparison between

$$\{Y_i^{obs} = 1 \mid S_i^{obs} = s, Z_i = 1\} \quad \text{and} \quad \{Y_i^{obs} = 1 \mid S_i^{obs} = s, Z_i = 0\} \qquad (ex.1)$$

If Z has any effect on S, then (ex.1) becomes (if Z is randomized)

$$\{Y_i(1) : S_i(1) = s\} \quad \text{and} \quad \{Y_i(0) : S_i(0) = s\}$$

and it is not a causal effect.

(3) Principal stratification.

(a) Definitions.

- **Definition 1.** A principal stratification is a partition of units by the joint post-treatment values $(S_i(0), S_i(1))$.

- **Definition 2.** A principal effect is a comparison between

$$\{Y_i(1) : S_i^p = s\} \quad \text{and} \quad \{Y_i(0) : S_i^p = s\}$$

where S_i^p indexes $(S_i(0), S_i(1))$.

(b) Main properties.

- **Property 1.** The stratum S_i^p is not effected by treatment Z.
- **Property 2.** A principal effect is a causal effect.

In many studies effects of interest are represented by principal effects.

(c) Case study. NEP (Frangakis et al. 2004)

3 Under-explored areas of interest.

(1) Censoring by death.

(2) Generalize results from one study to another (e.g., CDC anthrax vaccine trial, from macaques to human beings).

(3) Design issues.