SAMPLING DESIGN CONSIDERATIONS

• Some considerations in thinking about design.

• Hypothetical situations based on the CPUM.
THEORY IS THERE BUT...

• Theoretical framework supports a large class of (case and) control sampling designs.

• Nice, but so what?
BUT:

- What are useful study designs?

- What works and doesn’t work?
FACTORS TO CONSIDER IN DESIGNING A STUDY

• Study goals.

• Information needed in order to address the study goals.

• Information available on cohort members.

• More specifically, what are the costs associated with information components.
SOME STUDY DESIGN CONCEPTS

- Exposures - Factors of main interest.

- Confounders - Factors to control for. (Within-confounder comparisons increase validity.)

- Analytic variables - Factors needed to perform the key analyses, to be collected for the final (analytic) data set.

- Correlate or auxiliary factors - correlated with analytic variables (could be used to improve study design efficiency).
LOS ANGELES ENDOMETRIAL CANCER STUDY

- All cohort information to determine risk set membership, date of birth, date in, date out, marital stats computerized (cheap).

- Endometrial cancer occurrence information from the cancer registry (cheap).

- “Cheap” to determine risk set.

- Matching on year of birth and marital status at no additional cost, and improves validity.

Design: Simple random sample from risk sets matching on yob and marital status, medication histories just on case-control sample.
LOS ANGELES MELANOMA STUDY

Need detailed information from each subject.

- No information on cohort entire (NHW, <65) population of LA.

- Cancer registry provides source of cases (cheap).

- Address of case (auxiliary variable, cheap).

- Need to ascertain (matched) risk set membership (NHW and age/gender of case). (Needed for validity) (not too expensive)

- Extensive questionnaire, skin assessment (expensive).

Design: Use case-address to identify neighborhood and “walk” to identify control in the matched risk set.
STUDY OF ENDOMETRIAL HYPERPLASIA DIAGNOSIS AND ENDOMETRIAL CANCER

Need panel diagnosis of type of EH. (SH, CH, or AH types)

- Basic cohort information can be obtained from computer records (cheap).

- “Community” diagnosis available from paper records (not too expensive).

- AH quite rare but very predictive of EC.

Design: Two-phase study with batch quota sampling of risk set based on community diagnosis, then counter-matching within batch-quota sample.
USE OF COHORT INFORMATION

• If confounder (or confounder correlate) information is available for the cohort, may wish to use so that sampled controls are “closer” to the case with respect to this variable.

• If exposure (or exposure correlate) information is available for the cohort, may wish to use so that sampled controls are “farther” from the case (and each other).

• If at-risk is not known but correlate is available, may wish to use this information to reduce the number of subjects that must be “checked” before suitable controls are obtained.
HYPOTHETICAL SITUATIONS WITH THE CPUM

• Set up some situations.

• What are possible study designs.

• Restrict to case-control sampling type (we know how to analyze these).

• Compute the control selection probabilities to see how we would analyze the data.
HYPOTHETICAL SITUATION

• Goal is to estimate radon-lung cancer associations.

• CPUMs database has basic information for setting up the risk sets, date of birth info.

• Work/smoking history info on paper records and can be obtained through interview.

• Would like to have controls fairly closely matched to cases on year of birth.

5-year yob strata are a possibility, other designs?