The Cross-Sectional Study:
Investigating Prevalence and Association

Ronald A. Thisted

Departments of Health Studies and Statistics
The University of Chicago

CRTP Track I Seminar, Autumn, 2006
**Lecture Objectives**

1. Understand the structure of the cross-sectional study design,
2. Understand the advantages and disadvantages of this design, and
3. Understand the kinds of questions that can be addressed using cross-sectional data.
A Clinical Scenario: Are Kidney Stones and Hypertension Connected?

The Cross-Sectional Study
  The Logic
  The Structure
  Pros and Cons

Conducting a Cross-Sectional Study
  Steps in building a cross-sectional study
  Practical Issues

Discussion of clinical example
  The data set
  The questions
  Confounding and interaction

Data analysis issues

Suggested Reading
Outline

A Clinical Scenario: Are Kidney Stones and Hypertension Connected?

The Cross-Sectional Study
  The Logic
  The Structure
  Pros and Cons

Conducting a Cross-Sectional Study
  Steps in building a cross-sectional study
  Practical Issues

Discussion of clinical example
  The data set
  The questions
  Confounding and interaction

Data analysis issues

Suggested Reading
Are Kidney Stones and Hypertension Connected?

**Background**

- Women age 34–59 with history of kidney stones more likely to also have prior diagnosis of hypertension. (Madore, 1998)
- Men age 40–75 with stone history: same direction of association, but weaker. (Madore, 1998b)
- Multiple studies report relationship between BP and stones; wide variability in size of relationship.

**The Gap: What is really going on?**
Hypotheses:

▶ Some subgroups may be more susceptible to increased BP when stones are present than others.
▶ Heterogeneity of previous results may be due to differences in representation of high-risk subgroups.
▶ Association of BP and stones varies with (a) sex and (b) body size.

How to test these hypotheses?

▶ Need to be able to identify and study subgroups of interest.
▶ Need to avoid selection by stone status or blood pressure (so clinic-based samples are probably out)
▶ Would like to say something about subgroups as they exist in the adult population.

Investigated by Gillen, Coe, and Worcester using NHANES III: The Third National Health and Nutrition Examination Survey.
Hypotheses:

▶ Some subgroups may be more susceptible to increased BP when stones are present than others.
▶ Heterogeneity of previous results may be due to differences in representation of high-risk subgroups.
▶ Association of BP and stones varies with (a) sex and (b) body size.

How to test these hypotheses?

▶ Need to be able to identify and study subgroups of interest.
▶ Need to avoid selection by stone status or blood pressure (so clinic-based samples are probably out)
▶ Would like to say something about subgroups as they exist in the adult population.

Investigated by Gillen, Coe, and Worcester using NHANES III: The Third National Health and Nutrition Examination Survey.
Hypotheses:

- Some subgroups may be more susceptible to increased BP when stones are present than others.
- Heterogeneity of previous results may be due to differences in representation of high-risk subgroups.
- Association of BP and stones varies with (a) sex and (b) body size.

How to test these hypotheses?

- Need to be able to identify and study subgroups of interest.
- Need to avoid selection by stone status or blood pressure (so clinic-based samples are probably out).
- Would like to say something about subgroups as they exist in the adult population.

Investigated by Gillen, Coe, and Worcester using NHANES III: *The Third National Health and Nutrition Examination Survey.*
Outline

A Clinical Scenario: Are Kidney Stones and Hypertension Connected?

The Cross-Sectional Study
  The Logic
  The Structure
  Pros and Cons

Conducting a Cross-Sectional Study
  Steps in building a cross-sectional study
  Practical Issues

Discussion of clinical example
  The data set
  The questions
  Confounding and interaction

Data analysis issues

Suggested Reading
The Logic of Cross-Sectional Studies

- Looks at a “slice” of the population at a single point in time.
- If the selected sample is appropriately selected, composition of the sample reflects that in the population.
  - Simple random sample
  - Cluster sample
  - Stratified random samples
  - Multi-stage sample
- Perform pre-defined measurements and ascertainment.
  - Often include questionnaire/survey questions
- What can we do with this sample? We can estimate
  - Prevalence. What fraction of the population has a particular characteristic?
    [History of kidney stones? Diagnosed hypertension? SBP > 140?]
  - Association. What is the correlation between an “exposure” and an “outcome?”
    [Relationship of kidney stone history to HTN history]
The Structure of a Cross-Sectional Study

Begin

Study Population

Compare

[People with disease/outcome]

Risk Factor +

Risk Factor -

[People without disease/outcome]

Risk Factor +

Risk Factor -

Now
The Structure of a Cross-Sectional Study, Continued

Begin

Compare

[People with risk factor]

Outcome +

Outcome -

[People without risk factor]

Outcome +

Outcome -

Study Population

Now
Pros and Cons

▶ Advantages
  ▶ Cheaper/easier than longitudinal study: no follow-up required!
  ▶ Afford good control over the measurement/ascertainment process
  ▶ Can maximize completeness of key data (compared to retrospective study)
  ▶ Have greater control over precision of estimates in subgroups (stratified sampling)
  ▶ Often can be accomplished as secondary data analysis, that is, data collected by someone else (possibly for another purpose)
Pros and Cons, Continued

- **Disadvantages**
  - In secondary data analysis, no control over purpose, choice, or method of data collection
  - Cannot tell us about causal relationships (only correlation)
  - Generalizability limited by sampled population, population definition
  - Sample size requirements may be very large (especially when looking at rare outcomes or exposures)
  - Potential for selection bias.

*Example.* “Length-biased sampling” results from the fact that individuals with long courses of a disease are more likely to be the ones identified as prevalent cases than people with courses of short duration.
Outline

A Clinical Scenario: Are Kidney Stones and Hypertension Connected?

The Cross-Sectional Study
- The Logic
- The Structure
- Pros and Cons

Conducting a Cross-Sectional Study
- Steps in building a cross-sectional study
- Practical Issues

Discussion of clinical example
- The data set
- The questions
- Confounding and interaction

Data analysis issues

Suggested Reading
How to conduct a cross-sectional study

- Identify (and define) population of interest
  [Adults aged 17–90 with knowledge of stone hx].
- Define outcomes
  [Previous diagnosis of HTN; SBP].
- Define exposures (for correlational analysis)
  [Lifetime history of kidney stones]
- Create data collection “instruments”
  [Surveys, interviews, physical measurement procedures: SBP]
  Data collection forms are very useful for
  - Standardization
  - Protocol adherence
- Insure consistent ascertainment
  [Training of staff]
Sample from population appropriately
[Multistage sample of households].

Obtain consent, then “measure”

Analyze using appropriate statistical methods

May require special techniques to account for sampling design.

Statistical adjustment for confounders is usually necessary; we are (usually) after relationships that remain after adjusting for other factors

[Age, race, sex are differentially associated with both stone formation and blood pressure]
Sample from population appropriately [Multistage sample of households].

Obtain consent, then “measure”

Analyze using appropriate statistical methods

May require special techniques to account for sampling design.

Statistical adjustment for confounders is usually necessary; we are (usually) after relationships that remain after adjusting for other factors

[Age, race, sex are differentially associated with both stone formation and blood pressure]

OR . . .

Ask NORC.
Practical Issues for the Primary Cross-Sectionalist

- Non response
- Representativeness
- Logistics issues: cluster sampling, household contact
- Defining eligibility (target population)
- Defining measures in advance; respondent burden
- Data collection forms
Outline

A Clinical Scenario: Are Kidney Stones and Hypertension Connected?

The Cross-Sectional Study
   The Logic
   The Structure
   Pros and Cons

Conducting a Cross-Sectional Study
   Steps in building a cross-sectional study
   Practical Issues

Discussion of clinical example
   The data set
   The questions
   Confounding and interaction

Data analysis issues

Suggested Reading
Kidney Stones and BP

Gillen, *et al.*, used NHANES III data (publicly available) to address their research questions.

- Conducted 1988–1994 by NCHS
- National population-based sample
- Noninstitutionalized persons aged $>2$ months
- $n = 33,994$
- Stone history available on $n = 20,029$. 
Answering key questions

Have you ever had a kidney stone?
919 answer Yes (4.6%)
This is a simple prevalence calculation

BP outcomes:
1. Have you ever been told you have high blood pressure?
2. SBP, DBP, Pulse pressure

Relationship of sex to stone formation:
- Kidney stone Hx −: 54% women
- Kidney stone Hx +: 40% women  \( p < 0.001 \)
  Odds of stone history 1.73 higher in men than women

Relation of stone formation to hypertension:
“SFs were more likely to report a previous diagnosis of hypertension compared with non-SFs (32.7% vs 24.6%; \( P = 0.001 \)).”
[Univariate relationship]
Answering key questions

Have you ever had a kidney stone?
919 answer Yes (4.6%)
This is a simple prevalence calculation

BP outcomes:

1. Have you ever been told you have high blood pressure?
2. SBP, DBP, Pulse pressure

Relationship of sex to stone formation:

Kidney stone Hx −: 54% women
Kidney stone Hx +: 40% women \( p < 0.001 \)
Odds of stone history 1.73 higher in men than women

Relation of stone formation to hypertension:

“SFs were more likely to report a previous diagnosis of hypertension compared with non-SFs (32.7% vs 24.6%; \( P = 0.001 \)).”
[Univariate relationship]
Answering key questions

*Have you ever had a kidney stone?*
919 answer Yes (4.6%)
This is a simple *prevalence* calculation

*BP outcomes:*

1. *Have you ever been told you have high blood pressure?*
2. SBP, DBP, Pulse pressure

*Relationship of sex to stone formation:*

Kidney stone Hx −: 54% women
Kidney stone Hx +: 40% women $p < 0.001$
Odds of stone history 1.73 higher in men than women

*Relation of stone formation to hypertension:*

“SFs were more likely to report a previous diagnosis of hypertension compared with non-SFs (32.7% vs 24.6%; $P = 0.001$).”

[Univariate relationship]
Answering key questions

Have you ever had a kidney stone?
919 answer Yes (4.6%)
This is a simple prevalence calculation

BP outcomes:
1. Have you ever been told you have high blood pressure?
2. SBP, DBP, Pulse pressure

Relationship of sex to stone formation:
Kidney stone Hx −: 54% women
Kidney stone Hx +: 40% women $p < 0.001$
Odds of stone history 1.73 higher in men than women

Relation of stone formation to hypertension:
“SFs were more likely to report a previous diagnosis of hypertension compared with non-SFs (32.7% vs 24.6%; $P = 0.001$).”
[Univariate relationship]
Answering key questions

*Have you ever had a kidney stone?*
919 answer Yes (4.6%)
This is a simple prevalence calculation

*BP outcomes:*
1. *Have you ever been told you have high blood pressure?*
2. SBP, DBP, Pulse pressure

*Relationship of sex to stone formation:*
  - Kidney stone Hx −: 54% women
  - Kidney stone Hx +: 40% women \( p < 0.001 \)
    - Odds of stone history 1.73 higher in men than women

*Relation of stone formation to hypertension:*
  “SFs were more likely to report a previous diagnosis of hypertension compared with non-SFs (32.7% vs 24.6%; \( P = 0.001 \)).”
  [Univariate relationship]
Confounding: An Issue

- African-Americans less likely to form stones
- African-Americans more likely to have hypertension
- Want to “hold constant” AA effects when exploring stone-HTN effects
- Similarly for other confounders

Regression is one approach to confounders
Regression analysis

Table 2. Logistic Regression Results Modeling Self-Reported Diagnosis of Hypertension in the NHANES III Sample

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Adjusted Odds Ratio* (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>History of renal stones</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(yes v no)†</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>1.69 (1.33-2.17)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Men</td>
<td>1.20 (0.88-1.64)</td>
<td>0.237</td>
</tr>
<tr>
<td>Age (/5 y)</td>
<td>1.22 (1.20-1.24)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>African-American race</td>
<td>1.48 (1.29-1.69)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>BMI (/kg/m²)</td>
<td>1.10 (1.09-1.12)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Ever smoker (yes v no)</td>
<td>1.04 (0.89-1.22)</td>
<td>0.600</td>
</tr>
<tr>
<td>History of CVD (yes v no)</td>
<td>2.00 (1.66-2.41)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Diabetes (yes v no)</td>
<td>1.51 (1.25-1.82)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

*Adjusted for all covariates listed.
†P for interaction between history of renal stones and sex = 0.056.

What if degree of association depends on values of another variable?
Interaction effects

> SFs have higher BP than non-SFs
> This is stronger for heavier people
> This is more true for women than for men

Fig 1. Multiple linear regression estimates (and corresponding 95% CIs) of the average difference in SBP comparing SFs with non-SFs by sex and BMI quintile. All estimates are adjusted for age, race, BMI, smoking status, hypertension status, history of CVD, and diabetes.
Outline

A Clinical Scenario: Are Kidney Stones and Hypertension Connected?

The Cross-Sectional Study

  The Logic
  The Structure
  Pros and Cons

Conducting a Cross-Sectional Study

  Steps in building a cross-sectional study
  Practical Issues

Discussion of clinical example

  The data set
  The questions
  Confounding and interaction

Data analysis issues

Suggested Reading
Data Analysis/Statistics/Limitations

- Causality cannot be directly assessed
  - Do stones cause high BP, or does high BP predispose to stones?
  - Can view as “hypothesis generating” studies
  - Causality implies time course assessment
  - No experimental intervention

- Generalizability always an issue
  - Less so for population-based studies
  - Sampling method is key to generalizability
  - Relevance of defined population
    - [NHANES: noninstitutionalized population]
  - Particularly important for smaller-scale cross sections: “Cross section of what?”
  - Convenience samples are especially problem-prone since it is hard to know how people “select in” to the population: chart review, clinic waiting rooms, callers to a help line.

- Confounders
Data Analysis/Statistics/Limitations

- Causality cannot be directly assessed
  - Do stones cause high BP, or does high BP predispose to stones?
  - Can view as “hypothesis generating” studies
  - Causality implies time course assessment
  - No experimental intervention

- Generalizability always an issue
  - Less so for population-based studies
  - Sampling method is key to generalizability
  - Relevance of defined population
    - [NHANES: noninstitutionalized population]
  - Particularly important for smaller-scale cross sections: “Cross section of what?”
  - Convenience samples are especially problem-prone since it is hard to know how people “select in” to the population: chart review, clinic waiting rooms, callers to a help line.

- Confounders
Data Analysis/Statistics/Limitations

- Causality cannot be directly assessed
  - Do stones cause high BP, or does high BP predispose to stones?
  - Can view as “hypothesis generating” studies
  - Causality implies time course assessment
  - No experimental intervention

- Generalizability always an issue
  - Less so for population-based studies
  - Sampling method is key to generalizability
  - Relevance of defined population
    [NHANES: noninstitutionalized population]
  - Particularly important for smaller-scale cross sections: “Cross section of what?”
  - Convenience samples are especially problem-prone since it is hard to know how people “select in” to the population: chart review, clinic waiting rooms, callers to a help line.

- Confounders
Data Analysis/Statistics/Limitations, Continued

- **Confounders**
  - Statistical adjustment possible, *if* confounders are measured!
  - Regression methods are most common, although stratification can also be used
  - Unmeasured confounders
    - A major worry, but need to identify
    - Ingenuity may lead to suitable *proxies*
    - Unrecognized confounders = land mine
Outline

A Clinical Scenario: Are Kidney Stones and Hypertension Connected?

The Cross-Sectional Study
- The Logic
- The Structure
- Pros and Cons

Conducting a Cross-Sectional Study
- Steps in building a cross-sectional study
- Practical Issues

Discussion of clinical example
- The data set
- The questions
- Confounding and interaction

Data analysis issues

Suggested Reading
Suggested Reading

   An example of a community-based prevalence and correlation study.

   A useful paper cataloging important sources of bias.
   Available on the course website: http://health.bsd.uchicago.edu/thisted/epor/