Design, Conduct, and Analysis of Case Control Studies

Christy McKinney, PhD, MPH
Department of Oral Health Sciences
University of Washington
Learning Objectives

1. To learn when it is ideal to use a case control study design

2. To learn how selection bias can occur and some strategies to minimize it

3. To identify one aspect of the study management process that can improve data quality
Scope of Talk

Design
• Context
• Case and control selection and identification
• Data collection and measurement

Conduct
• Process overview
• Practical tips, tools and examples

Analysis
• General approach
Design
Level of Evidence

• Randomized controlled trial
• Prospective cohort study
• Case control, or retrospective cohort
• Cross-sectional
• Case series, case study, personal opinion
Viewpoints

• Limited Utility
  – Cannot determine causality
  – Need an randomized clinical trial

• Successes
  – Lung Cancer
  – Sudden Infant Death Syndrome
When to Use

• What is your question?
  – Detail study aims and each hypothesis

• Rare Disease
  – 1:100 or 10:100

• Efficient approach
  – Logistically and financially efficient
  – Statistically efficient
  – Odds ratio approximates relative risk
    • Biased if disease is not rare
Case Control Study Design

Cases

- Disease
  - Exposed
  - Not Exposed

Control

- No Disease
  - Exposed
  - Not Exposed
Design Overview

• Cases
  – All those with disease or outcome
  – Defined population

• Controls
  – Those without disease or outcome
  – Sample of source population that gave rise to cases

• Comparison
  – Measure exposure of interest
  – Compare proportion exposed in cases to controls
Thai Newborns
(Target)

Newborns in NE Thailand
(Source)

Hospital well-baby clinic

Hospital well-baby clinic
(Sample)

Hospital well-baby clinic

Cases

Enrolled Controls
Case Definition and Eligibility

• Clear, standardized, and discriminating
  – Defines the specific disease of interest
  – Defines source population
  – Defines eligibility to be in study

• Exclusions
  – Those without the disease
  – Those with other known causes (optional)
Example: Cases with Oral Cleft

Inclusion Criteria
• Child diagnosed with a non-syndromic cleft lip and/or palate at the Cleft Center: Cleft lip and palate (CL+P) or cleft lip only (CLO)
• Child has CL±P or CLO ± anomalies/conditions, not part of a syndrome
• Child <24 months of age
• Child born in northeast Thailand

Exclusion Criteria
• Child diagnosed with a cleft palate only (CPO) or atypical cleft (midline)
• Child diagnosed with a syndrome
• Biologic mother is not present
• Biologic mother does not speak Thai or Isaan dialect
Control Definition & Eligibility

Mirrors Case Definition
• But does not have disease
• Same source population
• Same eligibility criteria
• Same exclusion criteria
Example Comparison: Cases & Controls

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Case</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child has cleft lip with or without cleft palate</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Child has syndrome</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Child ± other congenital anomaly or condition</td>
<td>OK</td>
<td>OK</td>
</tr>
<tr>
<td>Child &lt;24 months</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Child born in NE Thailand</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Biologic mother not present</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Does not speak Thai or Isaan Dialect</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>
Case Control Study Design

Cases

Disease

Exposed

Not Exposed

Control

No Disease

Exposed

Not Exposed
Case Identification

• Cases have occurred when enrolled
  – Retrospective even with ongoing enrollment

• Case Subsets
  – Incident vs. Prevalent cases
    • Disease and characteristics for survival
    • Longer survival, slower disease process
    • May have different risk factors
  – Subset of cases (e.g. only one hospital)

• Potential Selection Bias
  – Misclassification of proportion exposed
Selection Bias: Cases

% low birth weight

- All cases
- Select subset
Control Identification

• Controls
  – *Represents source population*
  – *Population that gave rise to cases*
  – Sampling approaches
    • Random digit dialing
    • Lists: telephone, vital records, driver license, tax, health id numbers
    • Clinic or hospital based controls

• Potential Selection Bias
  – *Not represents source population*
Thai Newborns
(target)

% exposed in all newborns in Northeast Thailand (source population)

≠

% exposed in enrolled controls
Selection Bias: Controls

% low birth weight

Source population
Well-baby clinic
Hospitalized Infants
Matching Cases and Controls

What is matching?
• Process of selecting controls so similar to cases on key characteristics (e.g. age)

Why match?
• Reduces variability
• Improve ability to detect association

Why not match?
• Can be logistically complex
• Cannot examine as a risk factor
• You can adjust for the factor in your analysis
Number of Controls per Case

• Increases statistical efficiency (power)
  – Ability to detect associations

• 4:1 or 5:1 controls per case

• Cost determining factor
  – Expensive controls 1:1 is fine
  – Inexpensive controls can > 5:1
    • Enhance subgroup analysis
Data Collection

• Existing data
  – Data collected for general or other purposes
  – Limits determine by data available
  – Cost efficient

• Original data collection
  – Tailor data collection to questions of interest
  – More cost and labor intensive

• Depends on state of knowledge
Measurement

• Can examine multiple, unrelated exposures
  – Nutrition, supplement use
  – Genes

• Types of exposures
  – Biomarkers
  – Dental exam
  – Dental chart review
  – Self report (interview or self-administered)
  – Other
Measurement Details

• Timing of Measures
  – Timing of exposure
  – Time of disease onset

• Same time period
  – Prior to disease onset for cases at etiologic relevant time
  – Same prior period for controls

• Potential Bias
  – Does measure = pre-disease status?
  – Recall bias
Recall Bias

Self report data

Recall affects all subjects (non-differential)

• Length of time
• Clear specific time of interest

Differential recall

• Cases recall differently than controls
Conduct
Human Subjects Research Review
Training in Human Subjects Research

The CITI International Research Course Site

The CITI International Training Platform is a public access providing basic, foundational information for researchers, research staff and research ethics committee members involved in international research. This program now includes two learner tracks: Track 1 targets the international investigator (International Modules) while Track 2 is intended for non-US investigators, involved in U.S. Federally funded international research (US Federal Modules).

TRACK 1 The International Modules track is a new offering and the content was developed by Dr. Henry Silverman and his Fogarty International Center scholars in Egypt. The modules provide a general overview of the ethical issues central to conducting human subject research internationally. Translation of this track from English into other languages will begin in 2007.

TRACK 2 The US Federal Modules provide more detail about the US federally funded research requirements. US Federal Modules 1 through 3 have been translated into Simplified Chinese, French, Portuguese and Spanish. Modules 4 and 5 are available in English with some, but not all, of the content translated.

Topics for both tracks include:
1. History and Ethics
2. Research Review Processes
3. Informed Consent
4. International studies (Resource information/country specific information)
5. Course documents (Ethical guidance documents and links to research compliance information)

Although you will need to register to obtain a username and password for this course site, this program is available without charge.

To register for the IRB Training course please follow one of the two options below.

1. For new users, click here to register for the IRB Training course

2. If you are an existing user, log into the CITI Program website with your existing account and click on the "affiliate with another institution" link from your main menu. Select "IRB Training" from the list of participating institutions.

This site is provided as a service to the research community and is supported by CITI. The Fred Hutchinson Cancer Research Center and the University of Miami.

Questions, comments or suggestions about the International Training Platform or the CITI Program should be directed to the CITI Office at the University of Miami.

Thank you.
Protocol

• Often required part of IRB application
• Research proposal or independent document
• Specifies
  – Study design
  – Case definition and eligibility
  – Recruitment
  – Measurement: disease, exposure, covariates
  – Analysis plan
• Not day to day operations
Data Collection Instructions
Study Management Binder

- Contacts
- Protocol
- Manual of Operations
- Data Collection Instruments
- Laboratory Details
- Checklists
- Budget
Data Entry and Processing

The REDCap Consortium is composed of 566 active institutional partners from CTSA, GORC, RCMI and other institutions in 51 countries. The consortium supports a secure web application (REDCap) designed exclusively to support data capture for research studies.

The REDCap application allows users to build and manage online surveys and databases quickly and securely, and is currently in production use or development build status for more than 61,060 projects with over 86,000 users spanning numerous research focus areas across the consortium. To find out if your institution is already running REDCap, you will find contact information on the Consortium Partners page. Learn more about REDCap by watching a brief summary video (4 min).

Map of REDCap Consortium Partners

Recent publications using REDCap:
- Soluble ST2 as a Diagnostic and Prognostic Marker for Acute Heart Failure Syndromes. Open Biostat J 2012 Apr 20;2012(3):34-38.
Study Oversight

• Shared files with approved forms
  – Dropbox, Google docs
  – PDF versions best
  – Limited to forms, not study participant data

• Weekly recruitment tracking
  – Make IRB modification as needed

• Staff oversight
  – Training of staff
  – Spot check data collection
  – Questions, concerns, recommendations
Analysis
Statistical Approach

• Binary outcome
• Logistic regression methods
• Estimate odds ratios
• Odds ratios approximate relative risk
  – Only if disease is rare
Statistical Approach, cont.

• Matching
  – Adjust for matching factors
  – May need conditional logistic regression

• Resource
  – Logistic Regression: A Self-Learning Text
    (D.G. Kleinbaum and M. Klein)
Strengths & Limitations

Strengths
• Statistically efficient for rare diseases
• Logistical & financially practical for rare diseases
• Can examine multiple exposures

Limitations
• Odds ratio biased when disease is not rare
• Potential for bias: selection and recall
• Temporal sequence may be unclear
Thank you!

Questions?