When should you use a case study?

- When you can’t control the variables
- When there are many more variables than data points
- When you cannot separate phenomena from context
  - Phenomena that don’t occur in a lab setting
  - E.g. large scale, complex software projects
  - Effects can be wide-ranging.
  - Effects can take a long time to appear (weeks, months, years!)
- When the context is important
  - E.g. When you need to know how context affects the phenomena
- When you need to know whether your theory applies to a specific real world setting
Why conduct a case study?

- To gain a deep understanding of a phenomenon
  - Example: To understand the capability of a new tool
  - Example: To identify factors affecting communication in code inspections
  - Example: To characterize the process of coming up to speed on a project

- Objective of Investigation
  - Exploration- To find what’s out there
  - Characterization- To more fully describe
  - Validation- To find out whether a theory/hypothesis is true

- Subject of Investigation
  - An intervention, e.g. tool, technique, method, approach to design, implementation, or organizational structure
  - An existing thing or process, e.g. a team, releases, defects

Misuses of the term “Case Study”

- Not a case history
  - In medicine and law, patients or clients are “cases.” Hence sometimes they refer to a review of interesting instance(s) as a “case study”.

- Not an exemplar
  - Not a report of something interesting that was tried on a toy problem

- Not an experience report
  - Retrospective report on an experience (typically, industrial) with lessons learned

- Not a quasi-experiment with small n
  - Weaker form of experiment with a small sample size
  - Uses a different logic for designing the study and for generalizing from results
How can I tell it’s a case study?

- Has research questions set out from the beginning of the study
- Data is collected in a planned and consistent manner
- Inferences are made from the data to answer the research questions
- Produces an explanation, description, or causal analysis of a phenomenon
  - Can also be exploratory
- Threats to validity are addressed in a systematic way

Parts of a Case Study Research Design

- A research design is a “blueprint” for a study
  - Deals more with the logic of the study than the logistics
  - Plan for moving from questions to answers
  - Ensures data is collected and analyzed to produce an answer to the initial research question
  - (Analogy: research design is like a system design)

- Five parts of a case study research design
  1. Research questions
  2. Propositions (if any)
  3. Unit(s) of analysis
  4. Logic linking the data to the propositions
  5. Criteria for interpreting the findings
Part 1: Study Questions

- Study design always starts with research questions
  - Clarify precisely the nature of the research question
  - Ensure the questions can be answered with a case study
  - Generally, should be "how" and "why" questions.
  - Identify and interpret the relevant theoretical constructs

- Examples:
  - "Why do 2 organizations have a collaborative relationship?"
  - "Why do developers prefer this tool/model/notation?"
  - "How are inspections carried out in practice?"
  - "How does agile development work in practice?"
  - "Why do programmers fail to document their code?"
  - "How does software evolve over time?"
  - "Why have formal methods not been adopted widely for safety-critical software?"
  - "How does a company identify which software projects to start?"

Types of Case Studies

- Explanatory
  - Adjudicates between competing explanations (theories)
  - E.g. How important is implementation bias in requirements engineering?
    - Rival theories: existing architectures are useful for anchoring, vs. existing architectures are over-constraining during RE

- Descriptive
  - Describes sequence of events and underlying mechanisms
  - E.g. How does pair programming actually work?
  - E.g. How do software immigrants naturalize?

- Causal
  - Looks for causal relationship between concepts
  - E.g. How do requirements errors and programming errors affect safety in real time control systems?
    - See study by Robyn Lutz on the Voyager and Galileo spacecraft

- Exploratory
  - Used to build new theories where we don’t have any yet
  - Choose cases that meet particular criteria or parameters
  - E.g. Christopher Columbus’ voyage to the new world
  - E.g. What do CMM level 3 organizations have in common?
Part 2: Study Propositions

- Propositions are claims about the research question
  - State what you expect to show in the study
  - Direct attention to things that should be examined in the case study
  - E.g. "Organizations collaborate because they derive mutual benefits”

- Propositions will tell you where to look for relevant evidence
  - Example: Define and ascertain the specific benefits to each organization

- Note: exploratory studies might not have propositions
  - …but should lead to propositions for further study
  - …and should still have a clearly-stated purpose and clearly-stated criteria for success

- Analogy: hypotheses in controlled experiments

Part 3: Unit of Analysis

- Defines what a “case” is in the case study
  - Choice depends on the primary research questions
  - Choice affects decisions on data collection and analysis
  - Hard to change the unit of analysis once the study has started (but can be done if there are compelling reasons)
  - Note: good idea to use same unit of analysis as previous studies (why?)

- Often many choices:
  - E.g. for an exploratory study of extreme programming:
    - Unit of analysis = individual developer (case study focuses on a person’s participation in the project)
    - Unit of analysis = a team (case study focuses on team activities)
    - Unit of analysis = a decision (case study focuses on activities around that decision)
    - Unit of analysis = a process (e.g. case study examines how user stories are collected and prioritized)
Examples of Units of Analysis

- For a study of how software immigrants naturalize
  - Individuals?
  - … or the Development team?
  - … or the Organization?

- For a study of pair programming
  - Programming episodes?
  - … or Pairs of programmers?
  - … or the Development team?
  - … or the Organization?

- For a study of software evolution
  - A Modification report?
  - … or a File?
  - … or a System?
  - … or a Release?
  - … or a Stable release?

Why Defining your Unit of Analysis matters

- Clearly bounds the case study
  - …and tells you which data to collect

- Makes it easier to compare case studies
  - …incomparable unless you know the units of analysis are the same

- Avoid subjective judgment of scope:
  - e.g. disagreement about the beginning and end points of a process

- Avoids mistakes in inferences from the data
  - E.g. If your study proposition talks about team homogeneity…
  - …Won’t be able to say much if your units of analysis are individuals
**Part 4: Linking Logic**

- Logic or reasoning to link data to propositions
- One of the least well developed components in case studies
- Many ways to perform this
  - None as precisely defined as the treatment/subject approach used in controlled experiments
- One possibility is pattern matching
  - Describe several potential patterns, then compare the case study data to the patterns and see which one is closer

**Part 5: Interpretation Criteria**

- Criteria for interpreting a study’s findings
  - I.e. before you start, know how you will interpret your findings
- Also a relatively undeveloped component in case studies
  - Currently there is no general consensus on criteria for interpreting case study findings
  - [Compare with standard statistical tests for controlled experiments]
- Statistical vs. Analytical Generalization
  - Quantitative methods tend to sample over a population
  - Statistical tests then used to generalize to the whole population
  - Qualitative methods cannot use statistical generalization
  - Hence use analytical generalization
Generalization

**Statistical Generalization**
- First level generalization:
  - From sample to population
- Well understood and widely used in empirical studies
- Can only be used for quantifiable variables
- Based on random sampling:
  - Standard statistical tests tell you if results on a sample apply to the whole population
- Not useful for case studies
  - No random sampling
  - Rarely enough data points

**Analytical Generalization**
- Second level generalization:
  - From findings to theory
- Compares qualitative findings with the theory:
  - Does the data support or refute the theory?
  - Or: do they support this theory better than rival theories?
- Supports empirical induction:
  - Evidence builds if subsequent case studies also support the theory (& fail to support rival theories)
- More powerful than statistical techniques
  - Doesn’t rely on correlations
  - Examines underlying mechanisms

**Analytical and Statistical Generalization**

**LEVEL ONE**
- sample

**LEVEL TWO**
- theory
  - policy implication
- rival theory
  - rival policy implication

**SURVEY**
- population characteristics

**CASE STUDY**
- case study findings

**EXPERIMENT**
- experimental findings

**subjects**
How can I evaluate a case study?

Same criteria as for other empirical research:

- **Construct Validity**
  - Concepts being studied are operationalized and measured correctly

- **Internal Validity**
  - Establish a causal relationship and distinguish spurious relationships

- **External Validity**
  - Establish the domain to which a study’s findings can be generalized

- **Empirical Reliability**
  - Demonstrate that the study can be repeated with the same results

---

**Case Study Designs**

- **4 types of designs** (based on a 2x2 matrix)
  - Single-case vs. Multiple-case design
  - Holistic vs. Embedded design

---

Basic Types of Designs for Case Studies (Yin, page 40)
Holistic vs. Embedded Case Studies

- **Holistic** case study: Examines only the global nature of one unit of analysis (not any subunits)
  - E.g: a case study about an organization

- **Embedded** case study: Involves more than one unit of analysis by paying attention to subunit(s) within the case
  - E.g: a case study about a single organization may have conclusions about the people (subunits) within the organization

Holistic Designs

- **Strengths**
  - Convenient when no logical subunits can be defined
  - Good when the relevant theory underlying the case study is holistic in nature

- **Weaknesses**
  - Can lead to abstract studies with no clear measures or data
  - Harder to detect when the case study is shifting focus away from initial research questions
Embedded Designs

- **Strengths**
  - Introduces higher sensitivity to "slippage" from the original research questions

- **Weaknesses**
  - Can lead to focusing only on the subunit (i.e. a multiple-case study of the subunits) and failure to return to the larger unit of analysis

Rationale for Single-Case Designs

- As you might guess, a single-case design uses a single case study to address the research questions

- 5 reasons to use a single-case design
  - It represents the *critical* case in testing a well-formulated theory
    - The case meets all of the conditions for testing the theory thoroughly
  - It represents an *extreme or unique* case
    - Example: a case with a rare disorder
  - It is the *representative or typical* case, i.e. informs about common situations/experiences
    - Gain insights on commonplace situations
  - The case is *revelatory* – a unique opportunity to study something previously inaccessible to observation
    - Opens a new topic for exploration
  - The case is *longitudinal* – it studies the same case at several points in time
    - The corresponding theory should deal with the change of conditions over time
Multiple-Case Designs

- Useful when literal or theoretical replications provide valuable information

Advantages
- Evidence is considered more compelling
- Overall study is therefore regarded as more robust

Disadvantages
- Difficulty to find an appropriate number of relevant cases
- Can require extensive resources and time

Replication in Multiple-Case Studies

- Select each case so that it either:
  - Predicts similar results (*literal replication*)
  - Predicts contrasting results but for predictable reasons (*theoretical replication*)

- If all cases turn out as predicted, there is compelling support for the initial propositions
  - Otherwise the propositions must be revised and retested with another set of cases

- The theoretical framework of the study should guide the choices of replication cases
How Many Cases?

- How many literal replications?
  - It depends on the certainty you want to have about your results
  - Greater certainty with a larger number of cases
    - Just as with statistical significance measures
    - 2 or 3 may be sufficient if they address very different rival theories and the degree of certainty required is not high
    - 5, 6, or more may be needed for higher degree of certainty

- How many theoretical replications?
  - Consider the complexity of the domain under study
  - If you are uncertain whether external conditions will produce different results, you may want to include more cases that cover those conditions
  - Otherwise, a smaller number of theoretical replications may be used

Replication Logic vs. Sampling Logic

- Consider multiple-cases analogous to multiple experiments
  - Not analogous to multiple subjects in a single experiment!

- Replication logic (used in case studies) is different from sampling logic (used in surveys)
  - Sampling logic requires defining a pool of potential respondents, then selecting a subset using a statistical procedure
  - Responses from the subset are supposed to accurately reflect the responses of the entire pool

- Sampling logic does not fit with case studies
  - Case studies are not the best method for assessing the prevalence of phenomenon in a population
  - Case studies would have to cover both the phenomenon of interest and its context
    - Too many variables, which leads to way too many cases!
Replication Approach for Multiple-Case Studies

Multiple-Case Designs: Holistic or Embedded

- A multiple-case study can consist of multiple holistic or multiple embedded cases.
  - But there is no mixing of embedded and holistic cases in the same study.

- Note that for embedded studies, subunit data are not pooled across subunits.
  - Used to draw conclusions only for the subunit’s case.
Selecting Case Study Designs – Single or Multiple?

- If you have a choice and enough resources, multiple-case designs are preferred
  - Conclusions independently arising from several cases are more powerful
  - Differences in context of multiple cases with common conclusions improve the generalization of their findings
  - Capability to apply theoretical replications
- Single-case studies are often criticized due to fears about uniqueness surrounding the case
  - Criticisms may turn to skepticism about your ability to do empirical work beyond a single-case study
  - If you choose single-case design, be prepared to make an extremely strong argument justifying your choice for the case
- However, remember that in some situations single-case designs are the best (or only!) choice

Purposive Sampling of Cases

- Extreme or Deviant Case
  - E.g. outstanding success/notable failures, exotic events, crises.
- Intensity
  - Information-rich cases that clearly show the phenomenon (but not extreme)
- Maximum Variation
  - Choose a wide range of variation on dimensions of interest
- Homogeneous
  - Case with little internal variability - simplifies analysis
- Typical Case
  - Identify typical, normal, average case
- Stratified Purposeful
  - Identify subgroups and select candidates within each group
- Critical Case
  - If it’s true of this one case it’s likely to be true of all other cases.
- Snowball or Chain
  - Select cases that should lead to identification of further good cases
- Criterion
  - All cases that meet some criterion,
- Theory-Based
  - Manifestations of a theoretical construct
- Confirming or Disconfirming
  - Seek exceptions, variations on initial cases
- Opportunistic
  - Rare opportunity where access is normally hard/impossible
- Politically Important Cases
  - Attracts attention to the study
- Convenience
  - Cases that are easy/cheap to study (but means low credibility!)
- Or a combination of the above
Collecting the Evidence

Six Sources of Evidence
- Documentation
- Archival Records
- Interviews
- Direct Observation
- Participant-observation
- Physical Artifacts

Three Principles of Data Collection
- Use Multiple Sources of Evidence
- Create a Case Study Database
- Maintain a Chain of Evidence

Documentation
- Letters, memos, and other written communication
- Agendas, announcements, meeting minutes, reports of events
- Administrative documents
  - Proposals, progress reports, summaries and records
- Formal studies or evaluations of the same site
- Newspaper clippings, articles in media or newsletters
- Example: Classifying modification reports as adaptive, perfective or corrective based on documentation
Archival Records

- Service records
  - Clients served over a period of time
- Organizational records
  - Organizational charts and budgets
- Layouts
  - Maps and charts
- Lists of names and relevant articles
- Survey data
  - Census records
- Personal records
  - Diaries, calendars, telephone lists
- Example: Study of parallel changes to source code was based on revision control logs

Interviews

- Open-ended interviews
  - Address facts and opinions about an event
  - Flexible structure of interview (or no structure at all!)
- Focused interviews
  - Short period of time (about an hour)
  - Similar approach as open-ended.
- Formal surveys
  - Produce quantifiable data
- Example: Used semi-structured interviews to understand the effect of distance on coordination in teams
Direct Observation

- Field visits- creates opportunity for direct observation
- Photographs of site
  - Need permission in order to proceed!
- Can be used to calibrate self-reports
  - Example: Informal, impromptu interactions
- Example: Followed software developers around to characterize how they spend their time

Participant-observation

- Not a passive observer, but actually participate in setting
  - Employee of the company under study
- Provides an opportunity to understand the rationale and behavior of people and organization being studied
- Example: Seaman participated in 23 code inspections over period of five months at NASA/Goddard Space Flight Center’s Flight Dynamics Division
Physical Artifacts

- Technological tool, instrument, or device
- Artifacts can be collected or observed as part a field visit
- Works of art or types of physical evidence
- Example: Diachronic study of art records to determine whether right-handedness was a recent or old trait
  - Two rival hypotheses: Physiological predisposition vs Social/environmental pressure
  - Tested by counting unimanual tool usage in art representations
  - 1200 examples from 1500 BC to 1950, world sources
  - 92.6% used right hand
  - Geo/historical distribution uniformly high
  - Seems to support physiological interpretation that right-handedness is an age-old trait

Principles of Data Collection

- Use Multiple Sources of Evidence
- Create a Case Study Database
- Maintain a Chain of Evidence

These principles can be applied to all six data collection methods
Multiple Sources of Evidence

- Triangulation of data sources
- Basic idea: Collect evidence from more than one source pointing towards the same facts
  - Warning: Collecting data from several sources does not guarantee data triangulation!
- Example: Different approaches were used to collect data about how developers spend their time.
    - Collected cross-sectional and direct observation data
    - Collected longitudinal data

Convergence of Evidence (Figure 4.2)

- Documents
- Archival Records
- Observations (direct and participant)
- Structured Interviews and Surveys
- Open-ended Interviews
- Focus Interviews

FACT