

Foundations of Quantitative Social Science Research

Lecture 4-5: Research Design

Wanhong Huang

Spring 2025

Contents

1 Learning Objectives	4
2 Overview	4
3 Dimensions for Research Schemes	4
3.1 Intervention Structure	5
3.1.1 Experimental Intervention	5
3.1.2 Quasi-experimental Intervention	5
3.1.3 Observational Design	5
3.2 Assignment Mechanism	6
3.2.1 Randomized Assignment	6
3.2.2 Rule-based Assignment	6
3.2.3 Natural Assignment	6
3.2.4 Self-selection	6
3.3 Temporal Structure	7
3.3.1 Cross-sectional	7
3.3.2 Longitudinal	7
3.3.3 Time-series	7
3.4 Unit Tracking	7
3.4.1 Repeated Cross-sections	8
3.4.2 Panel Designs	8
3.4.3 Cohort Designs	8
3.5 Directionality of Inquiry	8
3.5.1 Prospective	8
3.5.2 Retrospective	8
3.5.3 Simultaneous	9
3.6 Comparison Logic	9
3.6.1 Between-unit Comparisons	9
3.6.2 Within-unit Comparisons	9
3.6.3 Historical Comparisons	9
3.6.4 External Comparisons	10
3.7 Data Provenance	10

3.7.1 Primary Data	10
3.7.2 Secondary Data	10
3.7.3 Mixed Sources	10
3.8 Sampling Strategy	10
3.8.1 Probability Sampling	11
3.8.2 Non-probability Sampling	11
3.8.3 Census	11
3.8.4 Selective Enrollment	11
3.9 Unit of Analysis and Observation	11
3.9.1 Individual-level	12
3.9.2 Aggregate-level	12
3.9.3 Multilevel	12
3.9.4 Ecological	12
3.10 Measurement Modality	12
3.10.1 Direct Observation	13
3.10.2 Self-report	13
3.10.3 Proxy Report	13
3.10.4 Administrative Records	13
3.10.5 Biomarkers and Physiological Measures	14
3.10.6 Digital Trace Data	14
3.10.7 Performance Tests	14
3.10.8 Mixed Measurement	14
3.11 Confounding Control Strategy	15
3.11.1 Randomization-based Control	15
3.11.2 Design-based Identification	15
3.11.3 Restriction and Stratification	16
3.11.4 Matching	17
3.11.5 Statistical Adjustment	17
3.11.6 Fixed Effects	17
3.11.7 Sensitivity Analysis	17
3.11.8 Negative Controls	18
3.11.9 No Explicit Control	18
3.12 Summary Table: Research Design Dimensions	18
3.13 Interdependencies Among Dimensions	18
4 Research Schemes Taxonomy	19
4.1 Randomized Controlled Trials	20
4.2 Quasi-Experimental Designs	21
4.3 Cohort Studies	21
4.4 Case-Control Studies	21
4.5 Cross-Sectional Studies	22
4.6 Panel Studies	22
4.7 Systematic Reviews and Meta-Analyses	22
4.8 Computational and Digital Trace Studies	22
4.9 Adaptive and Sequential Designs	23
4.10 Natural Experiments	23
4.11 Matching and Synthetic Control Designs	23
4.12 Regression Discontinuity Designs	24

4.13	Instrumental Variables Designs	24
4.14	Summary Table: Research Schemes and Dimensional Profiles	24
4.15	Summary Table: Measurement and Confounding Control by Scheme	24
5	Relationships Between Dimensions and Schemes	24
5.1	Design Space Navigation	26
5.2	Inferential Tradeoffs	26
6	Design Selection and Justification	27
6.1	Design Justification Template	31
6.2	Example: Design Justification for a Cohort Study of Educational Interventions	32
7	Recap: From Research Questions to Research Schemes	33
7.1	The Design Selection Workflow	33
7.1.1	Stage 1: Conceptual Clarification	33
7.1.2	Stage 2: Operationalization	34
7.1.3	Stage 3: Constraint Assessment	34
7.1.4	Stage 4: Dimensional Configuration	35
7.1.5	Stage 5: Scheme Recognition and Adaptation	36
7.2	Justifying Design Choices	37
7.3	Example: From Question to Design	37
7.4	Common Design Selection Patterns	39
7.5	Moving Forward: Design to Analysis	41
8	Case Studies: Research Schemes in Practice	41
8.1	Randomized Controlled Trials	41
8.1.1	Case Study 1: Moving to Opportunity Experiment	41
8.1.2	Case Study 2: Oregon Health Insurance Experiment	42
8.2	Quasi-Experimental Designs	43
8.2.1	Case Study 3: Minimum Wage and Employment (Card & Krueger)	43
8.2.2	Case Study 4: Class Size and Achievement (Angrist & Lavy, RDD)	44
8.2.3	Case Study 5: Military Service and Earnings (Angrist, IV Design)	45
8.3	Observational Designs	46
8.3.1	Case Study 6: Framingham Heart Study (Cohort)	46
8.3.2	Case Study 7: Nurses' Health Study (Cohort)	46
8.3.3	Case Study 8: Smoking and Lung Cancer (Case-Control)	47
8.3.4	Case Study 9: Social Capital and Health (Cross-Sectional)	48
8.4	Panel Studies	49
8.4.1	Case Study 10: Panel Study of Income Dynamics (PSID)	49
8.5	Digital Trace Studies	50
8.5.1	Case Study 11: Social Media and Political Polarization	50
8.5.2	Case Study 12: Emotional Contagion on Facebook	50
8.6	Meta-Analysis	51
8.6.1	Case Study 13: Early Childhood Education Effects (Meta-Analysis)	51
8.7	Synthesis: What Makes a Good Case Study?	52

1 Learning Objectives

By the end of this session, students will be able to:

1. Understand the foundational logic of quantitative social science research, including the full research lifecycle from question formulation to interpretation and dissemination.
2. Distinguish major research schemes (e.g., experimental, quasi-experimental, and observational designs) and explain how their structural dimensions constrain measurement, comparison, and inference.
3. Critically evaluate the identification assumptions underlying different quantitative methods, and assess their plausibility in applied research settings.
4. Select and justify appropriate measurement strategies and confounding control techniques for their own research questions.

2 Overview

Key Point

Quantitative social science research is not defined by individual methods or tools, but by coherent design choices that link research questions, data, methods, and assumptions across the entire research lifecycle.

This session introduces a unified framework for understanding quantitative social science research as an integrated process rather than a collection of isolated techniques. We examine how classical research designs (such as randomized experiments, cohort studies, and panel designs) and modern approaches (including digital trace analysis and adaptive designs) can be systematically compared along shared dimensions such as intervention structure, temporal organization, comparison logic, and confounding control.

Rather than focusing on algorithmic details, the emphasis is placed on methodological reasoning: how design choices encode assumptions, how measurement modalities shape inference, and how different research schemes achieve (or fail to achieve) credible causal and descriptive claims. This perspective provides students with a conceptual map that supports both critical reading of empirical studies and principled design of their own quantitative research projects.

3 Dimensions for Research Schemes

Research designs in quantitative social science can be systematically described using a finite set of *design dimensions*. Design dimensions are fundamental, conceptually distinct properties of a study that determine how data are generated, how comparisons are constructed, and what forms of inference are warranted. They are not statistical methods, estimators, or data collection techniques. Instead, dimensions function as *primitive building blocks* from which recognizable research schemes are constructed.

Each empirical study can be represented as a specific configuration of values along these dimensions. While the full Cartesian space of all possible combinations is large,

only a subset of combinations is logically coherent, practically feasible, and inferentially meaningful. Stable and recurrent configurations of dimensions give rise to canonical research schemes.

The principal dimensions are described below.

3.1 Intervention Structure

The intervention dimension characterizes whether and how the researcher actively manipulates exposure. This is perhaps the most fundamental design dimension, as it determines the degree of researcher control over the treatment assignment process and fundamentally shapes what types of causal inference are possible. In quantitative social science research, the intervention structure delimits the boundary between experimental and non-experimental inquiry, each with distinct epistemic strengths and practical constraints. The intervention dimension strongly constrains the scope of causal inference: experimental designs provide the strongest warrant for causal claims but face ethical and practical limitations, while observational designs offer greater external validity and feasibility at the cost of requiring stronger untestable assumptions.

3.1.1 Experimental Intervention

The researcher assigns treatment or exposure directly. Experimental designs afford the strongest claims to internal validity when properly implemented, as the researcher controls the allocation mechanism and can isolate the treatment effect from confounding influences through randomization or other design features. The ability to manipulate treatment allows the researcher to establish clear temporal precedence and to combine assignment with measurement protocols optimized for the research question.

3.1.2 Quasi-experimental Intervention

Treatment assignment is externally determined but structured in a way that enables causal comparison. The researcher does not control assignment but exploits systematic variation—such as policy changes, institutional rules, or natural discontinuities—that approximates experimental conditions under specific assumptions. Quasi-experimental designs bridge experimental and observational approaches, leveraging naturally occurring variation while maintaining some structural features that support credible causal inference.

3.1.3 Observational Design

No manipulation of exposure occurs. The researcher observes naturally occurring variation in treatment or exposure without intervening in the assignment process. Causal inference in observational settings requires stronger assumptions about the absence of confounding, typically operationalized through statistical adjustment, matching, or other methods that condition on measured covariates. Observational designs are essential when experimental manipulation is infeasible, unethical, or would fundamentally alter the phenomenon of interest.

3.2 Assignment Mechanism

The assignment mechanism specifies the process by which units receive treatments or exposures. While closely related to intervention structure, the assignment mechanism is analytically distinct: it describes *how* units come to be exposed, rather than *whether* the researcher controls exposure. This dimension is central to causal inference because the assignment mechanism determines what counterfactual comparisons are credible and what assumptions are required for valid causal identification. Different assignment mechanisms generate different patterns of confounding and selection, which in turn necessitate different identification strategies. Understanding the assignment mechanism is essential for assessing whether treatment and control groups are comparable and for determining what statistical or design-based methods can render them comparable.

3.2.1 Randomized Assignment

Allocation is determined by a stochastic process controlled by the researcher. Randomization balances observed and unobserved confounders in expectation, providing the foundation for unconfounded treatment effect estimation. The probability of assignment may be equal across units (simple randomization), vary by strata (stratified randomization), or depend on covariates (covariate-adaptive randomization). Regardless of the specific randomization scheme, the key property is that treatment assignment is statistically independent of potential outcomes.

3.2.2 Rule-based Assignment

Allocation follows deterministic rules or thresholds, such that treatment is assigned to all units meeting specified criteria (e.g., all units scoring above a cutoff, all individuals in certain geographic areas, or all applicants meeting eligibility requirements). Rule-based assignment enables identification strategies such as regression discontinuity designs when units near the threshold are comparable. The deterministic nature of assignment means that treatment status is a known function of observed characteristics, which can be exploited for causal inference under appropriate continuity assumptions.

3.2.3 Natural Assignment

Allocation is driven by external forces outside researcher control, such as policy implementation, geographic factors, historical events, or institutional processes. Natural assignment may approximate randomization if the assignment mechanism is plausibly exogenous to potential outcomes—that is, if factors determining treatment are unrelated to the outcome except through their effect on treatment receipt. Credibility of causal claims under natural assignment depends critically on substantive arguments about the assignment process and empirical assessment of balance on observed covariates.

3.2.4 Self-selection

Units select into exposure based on their own choices, preferences, or characteristics. Self-selection creates endogeneity concerns, as the factors influencing treatment choice (such as motivation, risk preferences, or anticipated benefits) may also directly affect outcomes. This generates confounding that cannot be eliminated by design but must be addressed through modeling assumptions. Controlling for selection-driven confounding requires

observing and adjusting for all relevant selection factors, an assumption (conditional independence or selection on observables) that is inherently untestable.

3.3 Temporal Structure

Temporal structure captures how observations are distributed over time. This dimension is fundamental to quantitative social science because many theoretical questions concern change, dynamics, and causal processes that unfold temporally. The temporal dimension determines whether a design can establish temporal precedence (a necessary condition for causal inference), track within-unit changes, detect time-varying effects, or distinguish short-term from long-term impacts. Different temporal structures enable different analytical strategies: cross-sectional designs efficiently estimate associations at a point in time; longitudinal designs track change and enable within-unit comparisons; time-series designs capture high-frequency dynamics and enable interrupted time-series analyses. The choice of temporal structure involves tradeoffs among resource requirements, attrition, time-varying confounding, and the ability to address temporally grounded research questions.

3.3.1 Cross-sectional

Units are observed at a single point in time. Cross-sectional designs are efficient and avoid attrition but cannot establish temporal precedence between exposure and outcome or distinguish cohort effects from age effects. All variation is between-unit variation, precluding within-unit comparisons.

3.3.2 Longitudinal

Units are observed at multiple discrete time points. Longitudinal designs enable analysis of change, assessment of temporal ordering, and estimation of both within-unit and between-unit effects. The interval between observations may range from days to years depending on the process under study.

3.3.3 Time-series

Units are observed at high-frequency or continuously over time. Time-series designs capture fine-grained temporal dynamics, enable detection of short-term fluctuations and long-term trends, and support methods such as interrupted time-series analysis that exploit detailed pre-intervention baselines.

3.4 Unit Tracking

Unit tracking describes whether the same observational units are followed over time. This dimension is distinct from temporal structure: a study may be longitudinal (multiple time points) but use repeated cross-sections (different units each time) or follow the same panel of units. Unit tracking determines what sources of variation are available for analysis and what types of unobserved heterogeneity can be controlled. Panel designs, which follow the same units, enable differencing out time-invariant unobserved confounders through fixed effects methods, but face challenges from attrition and the need to model time-varying confounding. Repeated cross-sections avoid attrition but cannot exploit

within-unit variation. Cohort designs balance some advantages of both by following a defined group without requiring full panel structure. This dimension critically affects identification strategies and the types of causal questions that can be addressed.

3.4.1 Repeated Cross-sections

Different units are observed at each time point. Repeated cross-sectional designs can track aggregate-level changes and cohort dynamics without the cost and attrition challenges of panel data, but cannot identify within-unit changes or control for unit-specific unobserved heterogeneity.

3.4.2 Panel Designs

The same units are repeatedly observed across all measurement occasions. Panel designs enable within-unit comparisons that control for stable unobserved characteristics, but require strategies to address selective attrition and time-varying confounding.

3.4.3 Cohort Designs

Units defined by a shared entry condition (e.g., birth year, program entry date, exposure to a common event) are followed over time. Cohort designs may involve complete panel tracking or accept some attrition, and naturally control for cohort-level confounders while allowing analysis of within-cohort heterogeneity.

3.5 Directionality of Inquiry

Directionality refers to the temporal relationship between exposure measurement and outcome observation. This dimension is crucial for quantitative social science because it determines vulnerability to specific types of bias—particularly recall bias, social desirability bias, and reverse causation—and affects the credibility of causal inference. Prospective designs measure exposure before outcome occurrence, establishing clear temporal precedence and minimizing recall distortion. Retrospective designs work backward from observed outcomes to assess prior exposures, which is efficient for rare outcomes but vulnerable to differential recall and selection bias. Simultaneous measurement precludes establishing temporal ordering altogether, limiting causal interpretation unless strong assumptions or external information justify claims about directionality. The choice of directionality involves tradeoffs between bias, cost, and feasibility.

3.5.1 Prospective

Exposure precedes outcome observation in both the causal process and the data collection sequence. Prospective designs establish clear temporal precedence, minimize recall bias in exposure measurement, and enable real-time data collection. However, they may require long follow-up periods and are inefficient for studying rare outcomes.

3.5.2 Retrospective

Outcome status determines exposure assessment: researchers identify individuals with the outcome of interest and then ascertain their prior exposure history. Retrospective designs are efficient for rare outcomes and can be implemented quickly, but are susceptible to

recall bias (differential memory of exposure by outcome status), selection bias (if outcome occurrence affects study inclusion), and survivor bias (if individuals must survive to outcome occurrence to be included).

3.5.3 Simultaneous

Exposure and outcome are measured concurrently, typically in cross-sectional surveys or assessments. Simultaneous measurement cannot establish temporal precedence from the design itself, making causal inference reliant on strong assumptions about the direction of causality, auxiliary information from theory or external studies, or structural modeling assumptions.

3.6 Comparison Logic

The comparison logic defines the source of the counterfactual—that is, how the design constructs an answer to the question "compared to what?" In causal inference, we seek to compare observed outcomes under treatment to counterfactual outcomes under control. Since the counterfactual is never directly observed, research designs use various comparison strategies to approximate it. The comparison logic determines what types of confounding the design can address and what assumptions are required. Between-unit comparisons assume different units are comparable (exchangeability); within-unit comparisons assume units are comparable to themselves at different times (temporal stability); historical comparisons assume the past predicts the counterfactual future (trend continuity); external comparisons assume other contexts are informative (transferability). Different comparison logics support different identification strategies and have characteristic vulnerabilities.

3.6.1 Between-unit Comparisons

Different units serve as comparisons for one another. Between-unit designs compare treated and untreated individuals, groups, or other entities. Validity requires that units are exchangeable—that is, comparable in terms of potential outcomes. This may be achieved through randomization, matching, statistical adjustment, or design-based strategies that render units comparable.

3.6.2 Within-unit Comparisons

Units serve as their own controls, comparing outcomes for the same unit at different times or under different conditions. Within-unit comparisons control for all time-invariant characteristics of the unit, observed and unobserved, but require assumptions about temporal stability and are vulnerable to time-varying confounding, carryover effects, and secular trends.

3.6.3 Historical Comparisons

Past observations act as counterfactuals for present or future outcomes. Historical comparisons, common in interrupted time-series and difference-in-differences designs, assume that pre-intervention trends would have continued absent intervention. Validity requires correct extrapolation of historical patterns into the counterfactual future.

3.6.4 External Comparisons

Other populations, contexts, or settings provide benchmarks for comparison (e.g., comparing one country to another, or study population outcomes to national norms). External comparisons rely on assumptions about the transferability of causal effects or the comparability of populations across contexts, assumptions that are often difficult to justify without strong theory or auxiliary evidence.

3.7 Data Provenance

Data provenance specifies the origin of the data used in analysis. This dimension affects data quality, construct validity, measurement error, completeness, and the alignment between measurement and theoretical constructs. Primary data are collected specifically for the research question at hand, allowing researchers to optimize measurement instruments, sampling, and timing for their inferential goals, but at substantial cost. Secondary data were generated for other purposes—administrative operations, prior research, commercial transactions—and may be available at low cost and large scale, but often involve compromises in construct validity and completeness. Understanding data provenance is essential for assessing measurement validity, anticipating missing data patterns, and evaluating whether available variables adequately operationalize theoretical constructs.

3.7.1 Primary Data

Collected specifically for the study. Primary data collection allows researchers to tailor instruments to theoretical constructs, optimize measurement timing, and implement quality control procedures. However, primary data collection is resource-intensive and may face challenges of recruitment, response rates, and attrition in longitudinal designs.

3.7.2 Secondary Data

Originally collected for other purposes, such as administrative operations, prior research studies, or commercial activities. Secondary data offer cost efficiency and often provide access to large samples or populations, but may involve construct validity concerns (available measures may not align well with theoretical constructs), missing data on key variables, and limited control over data quality.

3.7.3 Mixed Sources

Combination of primary and secondary data, such as linking survey responses to administrative records, augmenting administrative data with targeted primary data collection, or merging multiple secondary data sources. Mixed-source designs can capitalize on the strengths of each data type while potentially addressing their respective limitations, though linkage itself introduces challenges of matching accuracy and consent.

3.8 Sampling Strategy

Sampling strategy determines how units are selected from the target population into the study sample. This dimension governs external validity (the ability to generalize findings beyond the study sample) and the justification for statistical inference to broader populations. Probability sampling provides a foundation for design-based inference and

enables quantification of sampling uncertainty, but may be costly or infeasible for many populations. Non-probability sampling is more flexible and accessible but limits formal generalization to target populations. The sampling dimension interacts with other design features: population-representative samples enhance external validity but may complicate causal inference if they include substantial heterogeneity; convenience samples may enable cleaner causal estimates in homogeneous subgroups but raise questions about generalizability. Understanding sampling strategy is essential for interpreting study findings and assessing the populations to which results apply.

3.8.1 Probability Sampling

Units are selected with known, non-zero probability. Probability sampling methods include simple random sampling, stratified sampling (random sampling within strata defined by covariates), cluster sampling (random sampling of groups followed by observation of all or a sample of units within clusters), and multistage designs. Probability sampling supports design-based inference, allows calculation of sampling weights to adjust for unequal selection probabilities, and provides a basis for standard error estimation that accounts for the sampling design.

3.8.2 Non-probability Sampling

Units are selected without a probabilistic basis. Non-probability methods include convenience sampling (units that are readily accessible), purposive sampling (deliberate selection of information-rich cases), quota sampling (non-random selection to fill pre-determined quotas), and snowball sampling (chain-referral sampling). Non-probability samples may be suitable for exploratory research, hypothesis generation, or studies focusing on specific subpopulations, but do not support formal statistical inference to a defined population without strong assumptions.

3.8.3 Census

All units in the target population are included. Census designs eliminate sampling error and maximize statistical power, but are often feasible only for small, well-defined populations (e.g., all students in a school, all firms in an industry sector). Census data may still face coverage error if population enumeration is incomplete.

3.8.4 Selective Enrollment

Units meeting specific eligibility criteria are enrolled, often combining elements of purposive selection with probability or convenience sampling among eligible units. Selective enrollment is common in clinical trials (eligibility criteria ensure safety and relevance), evaluation studies (target program-eligible populations), and studies requiring specific characteristics for theoretical reasons. Selective enrollment affects generalizability to broader populations.

3.9 Unit of Analysis and Observation

This dimension distinguishes between the level at which data are collected (unit of observation) and the level at which inferences are made (unit of analysis). Misalignment

between these levels creates aggregation or disaggregation challenges and potential logical fallacies. This dimension is particularly important in quantitative social science because social phenomena are inherently multilevel: individuals nested in families, students in classrooms, workers in firms, citizens in nations. The choice of analytical level affects what questions can be addressed, what confounders must be considered, and what statistical methods are appropriate. Aggregate analyses may obscure individual-level heterogeneity; individual analyses may fail to capture contextual or compositional effects. Ecological fallacy (inferring individual-level relationships from aggregate data) and atomistic fallacy (ignoring higher-level structure) are key threats when observation and analysis units are misaligned.

3.9.1 Individual-level

Individuals serve as both observation and analysis units. Individual-level designs directly model individual outcomes and characteristics, providing maximum resolution for heterogeneity in treatment effects and outcomes. Standard regression methods and most causal inference techniques are designed for individual-level data.

3.9.2 Aggregate-level

Groups, organizations, or geographic units (e.g., schools, firms, neighborhoods, countries) are the units of both observation and analysis. Aggregate-level designs are appropriate when the theoretical question concerns collective entities or when individual data are unavailable. Aggregate analysis may be necessary for policy-relevant questions about organizational or jurisdictional effects.

3.9.3 Multilevel

Observations are nested within higher-level units, requiring hierarchical or multilevel modeling that explicitly represents this structure. Multilevel designs allow decomposition of variance across levels, estimation of cross-level interactions, and appropriate quantification of uncertainty that accounts for clustering. Multilevel analysis addresses questions about contextual effects and compositional effects while avoiding aggregation bias.

3.9.4 Ecological

Aggregate data are used to make inferences about aggregate-level processes or relationships. Ecological designs analyze associations among variables measured at the group level (e.g., relationships between county-level characteristics and county-level outcomes). Ecological inference is valid for understanding aggregate patterns but requires strong assumptions if used to infer individual-level relationships.

3.10 Measurement Modality

The measurement modality dimension characterizes how key variables are observed and recorded. This dimension has profound implications for measurement validity (whether measures capture intended constructs), reliability (consistency of measurement), and bias (systematic error in measurement). Different modalities involve distinct tradeoffs among

objectivity, construct validity, cost, participant burden, and ethical considerations. In quantitative social science, where many constructs of interest (attitudes, behaviors, social processes) are not directly observable, measurement modality choices fundamentally shape what can be studied and how credibly. The measurement dimension interacts with other design features: experimental designs often justify costly objective measures; observational studies often rely on self-report or administrative data; longitudinal designs must balance measurement burden against attrition. Understanding measurement error structure by modality is essential for assessing threats to validity and for selecting appropriate statistical methods.

3.10.1 Direct Observation

Researchers or trained observers record behaviors, events, or conditions in real time through systematic observation protocols. Examples include classroom observation schedules, workplace time-motion studies, structured field observations, and behavioral coding in laboratory settings. Direct observation minimizes recall bias and provides objective behavioral measurement, but is labor-intensive, may involve observer effects (reactivity), requires high inter-rater reliability, and is limited to observable behaviors rather than internal states or subjective experiences.

3.10.2 Self-report

Respondents provide information about themselves through surveys, interviews, or questionnaires. This includes standardized instruments (validated scales), open-ended interviews, ecological momentary assessment (repeated sampling of current states), and retrospective reports of past events or behaviors. Self-report provides access to internal experiences, subjective states, and private behaviors not accessible through observation, but is subject to recall bias, social desirability bias, acquiescence bias, satisficing, and question wording effects. The structure and quality of self-report measures vary widely, from rigorously validated psychometric instruments to ad hoc single-item questions.

3.10.3 Proxy Report

Information is obtained from knowledgeable informants rather than directly from the target individual. Examples include teachers reporting on student behavior, parents reporting on children's development, or family members reporting on patients' health status. Proxy reports may provide information when self-report is not feasible (young children, cognitively impaired individuals) or may triangulate with self-report to improve measurement. However, proxy reports may diverge from self-reports due to limited observability of internal states or private behaviors, differential interpretation of questions, or differing perspectives and incentives.

3.10.4 Administrative Records

Data are extracted from institutional or governmental databases created for operational purposes. Examples include health records (diagnoses, procedures, prescriptions), educational transcripts (grades, test scores, attendance), tax records (income, employment), criminal justice data (arrests, convictions, incarceration), and vital statistics (births,

deaths, marriages). Administrative data offer objectivity, large scale, population coverage, and longitudinal tracking, but may suffer from recording errors, systematic missingness related to system contact, construct validity challenges when operational categories do not align with research constructs, and changes in recording practices over time.

3.10.5 Biomarkers and Physiological Measures

Biological specimens or physiological sensors provide objective measures of physical states or biological processes. Examples include cortisol levels (stress), blood pressure, neuroimaging (brain structure and function), genetic markers, and cardiovascular or metabolic indicators. Biomarkers reduce self-report bias, provide objective indicators of physiological processes, and may capture information not accessible through self-report. However, they require specialized equipment and expertise, involve participant burden and potential health risks, may raise privacy concerns (especially genetic data), and often serve as imperfect proxies for constructs of theoretical interest (e.g., cortisol as a measure of stress).

3.10.6 Digital Trace Data

Passively collected behavioral data from digital platforms, sensors, or transaction systems. Examples include social media activity (posts, likes, network connections), GPS trajectories (mobility patterns), online purchases (consumer behavior), web browsing (information seeking), sensor networks (environmental exposures), and smartphone data (app usage, communication patterns). Digital traces offer high temporal resolution, ecological validity (naturalistic behavior), massive scale, and reduced participant burden (passive collection). However, they raise construct validity concerns (do observed behaviors map onto theoretical constructs?), privacy and ethical issues, selection bias (participation in platforms), algorithmic confounding (platform algorithms affect observed behavior), and heterogeneity in measurement across user populations.

3.10.7 Performance Tests

Standardized assessments or experimental tasks measure latent constructs through behavioral performance. Examples include cognitive ability tests (IQ, working memory), achievement tests (math, reading), physical fitness assessments, experimental economic games (risk preferences, cooperation), and implicit association tests (implicit attitudes). Performance tests provide objective behavioral measures of constructs that cannot be directly observed, often with established psychometric properties. However, they may be sensitive to testing conditions, require careful validation, and may not generalize beyond the testing context.

3.10.8 Mixed Measurement

Multiple measurement modalities are combined to triangulate constructs or validate measures. Examples include self-reported health combined with biomarkers and medical records, survey responses linked to administrative employment data, or direct observation combined with self-report interviews. Mixed measurement can capitalize on the strengths of different modalities, provide convergent validation, and enable assessment of measurement error through comparison across methods. However, mixed-mode designs

introduce complexities of data linkage, require managing discrepancies across measures, and may face challenges of consent for linking data sources.

The choice of measurement modality involves tradeoffs among validity, reliability, cost, participant burden, and ethical considerations. Objective measures (biomarkers, administrative records) reduce certain biases but may not capture subjective states or meanings. Self-report provides access to internal experiences but introduces response biases. Digital traces offer scale and naturalism but complicate construct interpretation. Measurement error structure varies systematically by modality: self-report tends toward systematic bias, administrative records toward random missingness, and digital traces toward measurement heterogeneity across subgroups.

3.11 Confounding Control Strategy

The confounding control dimension specifies how the design addresses potential confounding—the presence of common causes of treatment and outcome that threaten causal identification. Confounding is perhaps the central challenge in causal inference: if treatment and outcome share common causes, observed associations may reflect these shared causes rather than causal effects of treatment. The control strategy determines what assumptions must hold for valid causal inference and fundamentally shapes the credibility of causal claims. This dimension is crucial in quantitative social science because most social phenomena involve complex causal networks with multiple pathways connecting exposures to outcomes. Different control strategies embed identifying assumptions in different ways: randomization eliminates confounding by design; design-based identification exploits structural features to approximate randomization; statistical adjustment relies on conditioning on measured confounders. Understanding the confounding control dimension is essential for evaluating whether a study’s causal claims are justified and for determining what sensitivity analyses or robustness checks are needed.

3.11.1 Randomization-based Control

Random assignment of treatment eliminates confounding by ensuring treatment is statistically independent of all potential confounders, observed and unobserved. Randomization balances confounders in expectation, making treatment and control groups exchangeable and allowing causal effects to be estimated by simple mean differences (or regression adjustment to improve precision). This is the gold standard for internal validity because it requires no assumptions about which variables confound the relationship or how they operate. However, randomization is not always feasible (cannot randomize structural characteristics like race or gender), ethical (cannot randomize harmful exposures), or practical (cannot randomize many policy interventions). Even when feasible, randomization addresses only confounding, not other threats to validity such as attrition, measurement error, or spillover effects.

3.11.2 Design-based Identification

The research design exploits structural features that approximate randomization or render confounding implausible under explicit assumptions. Design-based approaches embed identifying assumptions in the design rather than in the statistical model, often providing more credible causal inference than pure statistical adjustment. Each design-based method relies on specific identifying assumptions:

Regression Discontinuity. Units just above and below a threshold are assumed comparable, isolating treatment effects at the cutoff. Identification requires continuity of potential outcomes at the threshold: individuals on either side differ only in treatment status, not in other characteristics affecting outcomes. RDD provides credible local causal effects but limited generalizability beyond the threshold region.

Instrumental Variables. An instrument affects the outcome only through its effect on treatment, breaking the confounding pathway. Identification requires instrument relevance (strong first stage), exclusion restriction (instrument affects outcome only through treatment), and independence (instrument uncorrelated with potential outcomes). IV estimates local average treatment effects for compliers—units whose treatment status is affected by the instrument.

Difference-in-Differences. Treatment and control groups share parallel counterfactual trends, allowing differencing to remove time-invariant confounding. Identification requires that treated and control groups would have followed parallel trends absent treatment. DiD controls for time-invariant confounding and common shocks but is vulnerable to time-varying confounding and violations of parallel trends.

Interrupted Time Series. Pre-intervention trends establish the counterfactual trajectory absent intervention. Identification requires that the pre-intervention trend would have continued absent intervention and that no concurrent events affect outcomes. ITS provides strong internal validity when long pre-intervention baselines establish stable trends, but is vulnerable to concurrent events and structural changes.

Natural Experiments. Exogenous shocks or policy changes create as-if random variation in exposure. Identification requires that the natural assignment mechanism is plausibly exogenous—that is, uncorrelated with potential outcomes except through treatment. Credibility depends on substantive arguments about the assignment process and empirical assessment of balance.

Each design-based approach relies on specific identifying assumptions that must be empirically assessed (through balance tests, placebo tests, or sensitivity analyses) and theoretically justified.

3.11.3 Restriction and Stratification

Analysis is restricted to units with similar values of known confounders, or conducted separately within strata defined by confounders. Restriction eliminates confounding by the restricted variables through design: if analysis includes only units of a given age, then age cannot confound the relationship. Similarly, stratified analysis ensures comparisons are made only within homogeneous strata. Restriction controls for observed confounders without requiring functional form assumptions, but does not address unmeasured confounders, reduces sample size and generalizability, and may introduce selection bias if restriction is based on post-treatment variables.

3.11.4 Matching

Treated and control units are paired or weighted to achieve balance on observed covariates. Matching methods include exact matching (pairing on identical covariate values), propensity score matching (pairing on predicted probability of treatment), coarsened exact matching (matching on coarsened covariates), and genetic matching (optimizing balance metrics). Matching controls for observed confounders by constructing a comparison group that resembles the treated group on measured characteristics. However, matching assumes no unobserved confounding (conditional independence given matched variables) and requires sufficient overlap (common support): for every treated unit there must exist comparable control units. Matching is often combined with regression adjustment (doubly robust estimation) to improve robustness.

3.11.5 Statistical Adjustment

Confounding is controlled through regression modeling, inverse probability weighting, or doubly robust estimation. Statistical adjustment conditions on observed confounders, estimating treatment effects within levels of confounders and then aggregating across levels. The validity of statistical adjustment depends on three assumptions: (1) conditional independence—all confounders are observed and included in the model; (2) correct functional form—the relationship between confounders, treatment, and outcome is correctly specified; and (3) positivity—all covariate combinations have non-zero probability of receiving each treatment level. Statistical adjustment is flexible and can accommodate continuous confounders and complex covariate patterns, but relies on strong untestable assumptions (especially conditional independence) and may be sensitive to model misspecification.

3.11.6 Fixed Effects

Within-unit or within-cluster variation is exploited to difference out time-invariant or unit-invariant confounders. Fixed effects models include unit-specific intercepts that absorb all stable characteristics of the unit, observed and unobserved. By comparing units to themselves over time (or clusters to themselves across observations), fixed effects control for all time-invariant confounding. Fixed effects are particularly powerful when important confounders are stable characteristics (ability, personality, neighborhood factors). However, fixed effects cannot identify effects of time-invariant treatments, are vulnerable to time-varying confounding, may exacerbate measurement error, and require correct specification of time-varying dynamics.

3.11.7 Sensitivity Analysis

Rather than eliminating confounding, sensitivity analyses quantify how strong unmeasured confounding would need to be to overturn the observed association. Sensitivity analysis methods include bounding approaches (calculating bounds on treatment effects under assumed confounding), tipping point analyses (determining the confounding strength required to nullify results), instrumental variable-based sensitivity tests, and simulated confounding scenarios. Sensitivity analysis does not resolve confounding but provides transparency about the robustness of causal claims to violations of identifying assumptions. Strong sensitivity analysis results (effects robust to substantial confounding) strengthen causal inference; weak results (effects eliminated by modest confounding) highlight the fragility of causal claims.

3.11.8 Negative Controls

Negative control outcomes (outcomes that should not be affected by treatment) or negative control exposures (exposures that should not affect the outcome) are used to detect residual confounding. If negative controls show associations with treatment or outcome, this suggests that confounding is present: the associations cannot be causal (by definition of negative controls) and therefore must reflect confounding or bias. Negative controls provide empirical evidence about the adequacy of confounding control without requiring explicit identification of confounders. However, specifying valid negative controls requires strong substantive assumptions about what should and should not be causally related.

3.11.9 No Explicit Control

Some observational studies make no systematic attempt to control confounding, either because the research question is purely descriptive (not causal) or because causal interpretation relies on conditional independence assumptions implicit in the research question. Studies without explicit confounding control may still provide valuable descriptive evidence, generate hypotheses, or inform causal inference when combined with theory or evidence from other sources. However, causal claims from such studies are generally weak and require strong justification.

The choice of confounding control strategy is central to causal inference. Randomization-based and design-based strategies embed identifying assumptions in the research design itself, while statistical adjustment and matching embed assumptions in the analysis model. Transparent articulation of the chosen strategy and its underlying assumptions is essential for evaluating the credibility of causal claims. Multiple strategies may be combined in a single study. For example, a difference-in-differences design may additionally employ matching to improve covariate balance or regression adjustment to control for time-varying confounders. Combining strategies can provide robustness checks and strengthen causal inference if results are consistent across methods with different assumptions.

Multiple strategies may be combined in a single study. For example, a difference-in-differences design may additionally employ matching to improve covariate balance or regression adjustment to control for time-varying confounders.

3.12 Summary Table: Research Design Dimensions

3.13 Interdependencies Among Dimensions

Design dimensions are not fully independent. Certain dimensional values constrain or necessitate others, creating structural dependencies that shape the feasible design space.

Logical dependencies: Some combinations are logically impossible. Randomized assignment requires experimental or quasi-experimental intervention structure; retrospective directionality is incompatible with randomization; within-unit comparisons require longitudinal temporal structure.

Practical dependencies: Certain combinations, while not logically impossible, are rarely implemented due to practical constraints. For example, randomized experiments typically use primary data collection; meta-analyses necessarily use secondary data; digital trace studies typically employ observational designs.

Inferential dependencies: The choice along one dimension affects which values are appropriate on other dimensions for valid inference. High-frequency time-series data

Table 1: Taxonomy of Research Design Dimensions

Dimension	Possible Values
Intervention Structure	Experimental
	Quasi-experimental
	Observational
Assignment Mechanism	Randomized
	Rule-based
	Natural
	Self-selection
Temporal Structure	Cross-sectional
	Longitudinal
	Time-series
Unit Tracking	Repeated cross-sections
	Panel
	Cohort
Directionality of Inquiry	Prospective
	Retrospective
	Simultaneous

may enable design-based identification strategies (e.g., interrupted time series) that are infeasible with cross-sectional data. Panel data enable fixed effects control that cross-sectional designs cannot support.

Measurement-design dependencies: Certain measurement modalities align naturally with particular design structures. Biomarkers and performance tests are common in experimental designs where precise measurement justifies the cost. Digital trace data typically appear in observational designs with natural assignment. Administrative records often enable large-scale quasi-experimental designs exploiting policy variation.

Control strategy dependencies: The feasibility and appropriateness of confounding control strategies depend on other dimensional values. Randomization-based control defines experimental intervention structure. Design-based identification requires specific temporal structures (longitudinal for DiD, cross-sectional with threshold for RDD). Fixed effects require panel data. Matching and statistical adjustment require rich covariate measurement.

Understanding these interdependencies helps researchers navigate the design space efficiently and avoid internally inconsistent configurations.

4 Research Schemes Taxonomy

Research schemes are coherent, recognizable configurations of design dimensions that support specific inferential goals. Unlike dimensions, which are abstract descriptors, research schemes function as *inferential templates*. Each scheme is associated with characteristic assumptions, strengths, and limitations.

This section presents a taxonomy of commonly used research schemes in quantitative social science.

Table 2: Taxonomy of Research Design Dimensions (Continued)

Dimension	Possible Values
Comparison Logic	Between-unit Within-unit Historical External
Data Provenance	Primary Secondary Mixed sources
Sampling Strategy	Probability Non-probability Census Selective enrollment
Unit of Analysis	Individual-level Aggregate-level Multilevel Ecological
Measurement Modality	Direct observation Self-report Proxy report Administrative records Biomarkers Digital trace data Performance tests Mixed measurement
Confounding Control	Randomization-based Design-based identification Restriction / stratification Matching Statistical adjustment Fixed effects Sensitivity analysis Negative controls No explicit control

4.1 Randomized Controlled Trials

Randomized controlled trials combine experimental intervention with randomized assignment and prospective outcome measurement. Comparisons are typically between treatment and control groups, though within-unit designs (e.g., crossover trials) may also be employed.

RCTs provide strong internal validity under correct implementation, but may face constraints related to ethics, feasibility, and external validity. Common variants include parallel group designs, cluster-randomized trials, factorial designs, and adaptive randomization schemes.

Dimensional configuration: Experimental intervention, randomized assignment, prospective directionality, between-unit (or within-unit) comparison, primary data collection, randomization-based confounding control, typically direct observation or performance tests for outcomes.

Key identifying assumption: Random assignment ensures treatment independence from potential outcomes (unconfoundedness by design).

4.2 Quasi-Experimental Designs

Quasi-experimental designs exploit structured variation in exposure without full randomization. Common examples include difference-in-differences, regression discontinuity designs, instrumental variables, and interrupted time-series analyses.

These designs rely on design-specific assumptions, such as parallel trends or continuity at thresholds, to support causal inference. Each variant has characteristic data requirements and identifying assumptions.

Dimensional configuration: Quasi-experimental intervention, rule-based or natural assignment, longitudinal or cross-sectional temporal structure, between-unit and/or within-unit comparison, primary or secondary data, design-based identification for confounding control.

Key identifying assumptions: Vary by specific design (e.g., parallel trends for DiD, continuity at threshold for RDD, exclusion restriction for IV).

4.3 Cohort Studies

Cohort studies follow a group of units defined by a shared initial condition over time, typically in an observational setting. Exposure is measured prior to outcome observation, enabling prospective analysis of risk and incidence.

While cohort designs support temporal ordering, they remain vulnerable to confounding. Cohort studies may employ restriction, statistical adjustment, or matching to control confounding.

Dimensional configuration: Observational, natural or self-selection assignment, longitudinal temporal structure, cohort tracking, prospective directionality, between-unit comparison, primary data, statistical adjustment or restriction for confounding control, typically self-report or biomarkers for exposure and outcomes.

Key identifying assumption: Conditional independence given measured covariates (no unmeasured confounding).

4.4 Case–Control Studies

Case–control studies select units based on outcome status and retrospectively assess exposure. Comparisons are made between cases and controls to identify associations with the outcome of interest.

These designs are efficient for rare outcomes but are susceptible to selection and recall bias. Case-control studies typically use matching on key confounders and statistical adjustment in analysis.

Dimensional configuration: Observational, selective enrollment based on outcome status, cross-sectional or retrospective temporal structure, retrospective directionality, between-unit comparison, primary or secondary data, matching and statistical adjustment for confounding control, often self-report or administrative records for exposure.

Key identifying assumptions: Controls represent the source population that gave rise to cases; exposure measurement is unbiased by outcome status; conditional independence given covariates.

4.5 Cross-Sectional Studies

Cross-sectional studies observe units at a single point in time, measuring exposures and outcomes simultaneously. They are primarily descriptive and correlational in nature.

Causal interpretation is generally limited without additional assumptions or external information. When causal claims are made, they typically rely on statistical adjustment or instrumental variables.

Dimensional configuration: Observational, self-selection or natural assignment, cross-sectional temporal structure, simultaneous directionality, between-unit comparison, primary or secondary data, statistical adjustment if causal inference attempted, self-report or administrative records common.

Key limitation: Cannot establish temporal precedence; vulnerable to reverse causation.

4.6 Panel Studies

Panel studies repeatedly observe the same units over time without experimental intervention. By combining within-unit and between-unit variation, panel designs enable analysis of change and dynamic relationships.

Attrition and time-varying confounding pose central challenges. Panel designs often employ fixed effects to control time-invariant confounding, with additional adjustment for time-varying confounders.

Dimensional configuration: Observational, self-selection or natural assignment, longitudinal temporal structure, panel tracking, prospective directionality, within-unit and between-unit comparison, primary or secondary data, fixed effects and statistical adjustment for confounding control, varied measurement modalities.

Key identifying assumption: No time-varying unmeasured confounding (for fixed effects); correct dynamic specification.

4.7 Systematic Reviews and Meta-Analyses

These schemes operate on studies rather than individuals as units of analysis. By synthesizing evidence across multiple studies, they aim to estimate aggregate effects or summarize bodies of evidence.

Inference depends critically on study quality and comparability. Meta-analyses use statistical methods to combine effect estimates, accounting for sampling variability and between-study heterogeneity.

Dimensional configuration: Observational (of published studies), secondary data, between-study comparison, study-level unit of analysis, no primary confounding control (but may assess study quality), measurement via published effect estimates.

Key threats: Publication bias, selective reporting, incomparability of study designs and populations.

4.8 Computational and Digital Trace Studies

These designs analyze passively collected, high-frequency data such as digital traces or sensor records. They are typically observational and emphasize pattern detection and behavioral dynamics.

Construct validity and ethical considerations are central concerns. Large sample sizes enable detection of small effects but do not resolve confounding.

Dimensional configuration: Observational, natural assignment, time-series or longitudinal temporal structure, panel or repeated cross-sections tracking, secondary data (digital trace measurement), within-unit and between-unit comparison, statistical adjustment or design-based methods for confounding control.

Key challenges: Construct validity (do digital traces measure intended constructs?); selection into platform usage; algorithmic confounding.

4.9 Adaptive and Sequential Designs

Adaptive designs allow design features, such as treatment assignment probabilities, to evolve in response to accumulating data. Examples include sequential trials and SMART (Sequential Multiple Assignment Randomized Trials) designs.

These schemes increase efficiency but require careful integration of design and analysis to maintain Type I error control and valid inference.

Dimensional configuration: Experimental, adaptive randomized assignment, longitudinal temporal structure, panel tracking, prospective directionality, between-unit and within-unit comparison, primary data, randomization-based confounding control, direct observation or performance tests.

Key requirement: Pre-specified adaptation rules to preserve inferential validity.

4.10 Natural Experiments

Natural experiments exploit exogenous variation in exposure created by events, policies, or institutional features outside researcher control. The exposure variation approximates random assignment conditional on appropriate controls.

Credibility depends on the plausibility of the as-if random assignment assumption and the absence of concurrent confounding events. Common contexts include policy changes, lottery-based allocation, geographic discontinuities, and timing-based variation.

Dimensional configuration: Quasi-experimental, natural assignment, longitudinal or cross-sectional temporal structure, between-unit comparison, prospective or retrospective directionality, secondary or mixed data sources, design-based identification (exploiting exogenous variation).

Key identifying assumption: Assignment mechanism is as-if random conditional on observables; no concurrent confounders.

4.11 Matching and Synthetic Control Designs

These designs construct comparison groups by identifying or creating units that resemble treated units on pre-treatment characteristics. Matching can be exact or based on propensity scores; synthetic controls weight donor units to replicate pre-treatment trajectories.

Identifying valid counterfactuals requires strong overlap and no unobserved confounding assumptions. Matching designs are often combined with statistical adjustment (doubly robust estimation).

Dimensional configuration: Observational or quasi-experimental, self-selection or natural assignment, longitudinal or cross-sectional temporal structure, between-unit com-

parison via constructed controls, secondary or primary data, matching-based confounding control with potential statistical adjustment, varied measurement modalities.

Key identifying assumptions: Conditional independence given matched covariates; common support (overlap); stable unit treatment value assumption (SUTVA).

4.12 Regression Discontinuity Designs

Regression discontinuity designs (RDD) exploit treatment assignment rules based on whether a continuous assignment variable crosses a known threshold. Units just above and below the threshold are assumed comparable, enabling causal inference at the cutoff.

RDDs provide credible local causal estimates but have limited external validity beyond the threshold region.

Dimensional configuration: Quasi-experimental, rule-based assignment at threshold, cross-sectional or longitudinal temporal structure, between-unit comparison in narrow bandwidth around threshold, primary or secondary data, design-based identification exploiting discontinuity, varied measurement modalities.

Key identifying assumptions: Continuity of potential outcomes at threshold; no manipulation of assignment variable; correct functional form specification.

4.13 Instrumental Variables Designs

Instrumental variables (IV) designs use an instrument—a variable that affects treatment but not the outcome except through treatment—to isolate exogenous variation in treatment. IV estimates the local average treatment effect for compliers (units whose treatment status is affected by the instrument).

IV designs require strong substantive justification for instrument validity and are sensitive to weak instruments.

Dimensional configuration: Quasi-experimental or observational, natural or rule-based assignment of instrument, cross-sectional or longitudinal temporal structure, between-unit comparison, primary or secondary data, design-based identification via instrumental variables, varied measurement modalities.

Key identifying assumptions: Instrument relevance (strong first stage); exclusion restriction (instrument affects outcome only through treatment); monotonicity (no defiers); independence of instrument from potential outcomes.

4.14 Summary Table: Research Schemes and Dimensional Profiles

4.15 Summary Table: Measurement and Confounding Control by Scheme

5 Relationships Between Dimensions and Schemes

The relationship between design dimensions and research schemes is not one-to-one. A single scheme may admit variation along certain dimensions while remaining fixed on others. For example, randomized controlled trials are necessarily experimental and randomized, but may vary in their temporal structure (single follow-up versus repeated measurements) or comparison logic (parallel groups versus crossover designs).

Table 3: Canonical Research Schemes and Their Dimensional Profiles

Scheme	Intervention	Assignment	Temporal	Direction	Comparison	Confounding Control
RCT	Experimental	Randomized	Longitudinal	Prospective	Between Within	Randomization
Quasi-Exp.	Quasi-exp.	Rule Natural	Varied	Prospective	Between Within	Design-based
Cohort Study	Observational	Natural Self	Longitudinal	Prospective	Between	Statistical adj.
Case-Control	Observational	Outcome-based	Cross-sect.	Retrospective	Between	Matching + adj.
Cross-Sectional	Observational	Natural Self	Cross-sect.	Simultaneous	Between	Statistical adj.
Panel Study	Observational	Natural Self	Longitudinal	Prospective	Within Between	Fixed effects + adj.
Meta-Analysis	Observational	N/A (studies)	N/A	Retrospective	Between-study	None (quality assess.)
Digital Trace	Observational	Natural	Time-series	Prospective	Within Between	Statistical/design
Adaptive Design	Experimental	Adaptive Random	Longitudinal	Prospective	Between Within	Randomization
Natural Exp.	Quasi-exp.	Natural	Varied	Varied	Between	Design-based
Matching/Synth	Obs. Quasi	Self Natural	Longitudinal	Prospective	Between (matched)	Matching
RDD	Quasi-exp.	Rule-based	Cross-sect. Long.	Prospective	Between (local)	Design-based (disc.)
IV Design	Quasi Obs.	Natural Rule (IV)	Varied	Varied	Between	Design-based (IV)

Table 4: Measurement Modalities and Confounding Control by Research Scheme

Scheme	Measurement Modalities	Confounding Control	Key Assumption
RCT	Direct observation Performance tests Biomarkers Self-report	Randomization-based	Random assignment
Quasi-Experimental	Administrative records Self-report Mixed sources	Design-based (DiD, RDD, IV, ITS)	Design-specific validity
Cohort Study	Self-report Biomarkers Administrative records	Adjustment Restriction Matching	Conditional independence
Case-Control	Self-report Administrative records Proxy report	Matching + adjustment	No differential recall bias
Cross-Sectional	Self-report Administrative records	Adjustment; IV if available	Causal sufficiency (if causal)
Panel Study	Self-report Administrative records Mixed	Fixed effects + adjustment	No time-varying confounding
Meta-Analysis	Published effect estimates	Quality assessment	Study comparability
Digital Trace	Digital trace data Sensor data	Adjustment; design-based	Construct validity
Adaptive Design	Direct observation Performance tests Biomarkers	Adaptive randomization	Valid adaptation rules
Natural Experiment	Administrative records Mixed sources	Exogenous variation	As-if random assignment
RDD	Administrative records Self-report Mixed	Discontinuity design	Continuity at cutoff
IV Design	Administrative records Self-report Mixed	Instrumental variables	Exclusion restriction

Conversely, multiple schemes may share the same configuration on some dimensions while differing on others. Both cohort studies and panel studies are observational and longitudinal, but panel studies emphasize within-unit variation whereas traditional cohort studies focus on between-unit comparisons.

This many-to-many mapping reflects the fact that dimensions describe *primitive features*, while schemes represent *coherent patterns* that have crystallized through disciplinary practice. Understanding both levels of description enables researchers to:

1. Recognize the abstract structure underlying familiar designs.
2. Adapt canonical schemes to novel contexts by modifying specific dimensions.
3. Identify hybrid or emerging designs that combine features from multiple traditions.
4. Communicate design choices precisely by specifying dimensional configurations.

5.1 Design Space Navigation

The dimensional framework suggests that research design can be understood as navigation through a constrained design space. Not all combinations of dimensional values are feasible: randomized assignment is incompatible with retrospective directionality; within-unit comparisons require longitudinal data; meta-analyses necessarily use secondary data; design-based confounding control strategies require specific temporal and assignment structures.

These logical constraints reduce the effective dimensionality of the design space. Additionally, practical, ethical, and resource constraints further limit feasible designs for any given research question. The researcher's task is to identify the configuration that maximizes inferential strength given these constraints.

5.2 Inferential Tradeoffs

Different dimensional configurations embody different inferential tradeoffs. Experimental intervention and randomized assignment maximize internal validity but may compromise external validity, feasibility, or ethical acceptability. Observational designs offer greater external validity and practical feasibility but require stronger modeling assumptions to support causal claims.

Similarly, longitudinal designs enable analysis of temporal dynamics and change but face challenges related to attrition, time-varying confounding, and resource demands. Cross-sectional designs are efficient but cannot establish temporal precedence.

Measurement modality choices involve tradeoffs between objectivity and construct validity. Administrative records and biomarkers reduce self-report bias but may not capture the constructs of theoretical interest. Self-report provides access to subjective states but introduces multiple response biases. Digital trace data offer naturalism and scale but complicate construct interpretation and raise privacy concerns.

Confounding control strategies similarly involve tradeoffs. Randomization eliminates confounding but may be infeasible or unethical. Design-based identification provides credible causal inference under weaker assumptions than statistical adjustment but requires specific data structures and is often limited to local or context-specific effects. Statistical adjustment is flexible but relies on untestable conditional independence assumptions.

Understanding these tradeoffs requires recognizing that no single design dominates across all criteria. The optimal design depends on the research question, the population and setting, available resources, and the relative importance of different inferential goals.

Design Dimensions

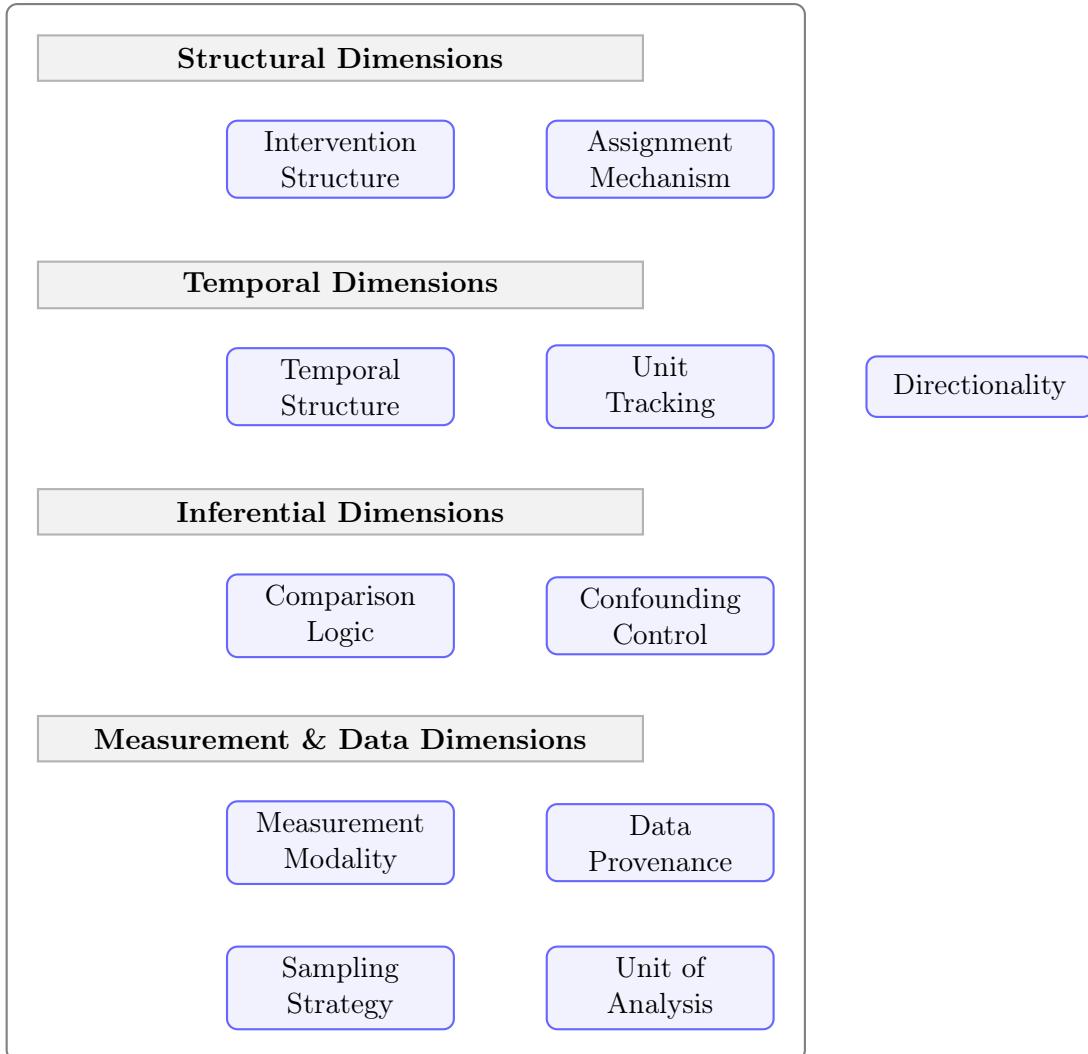
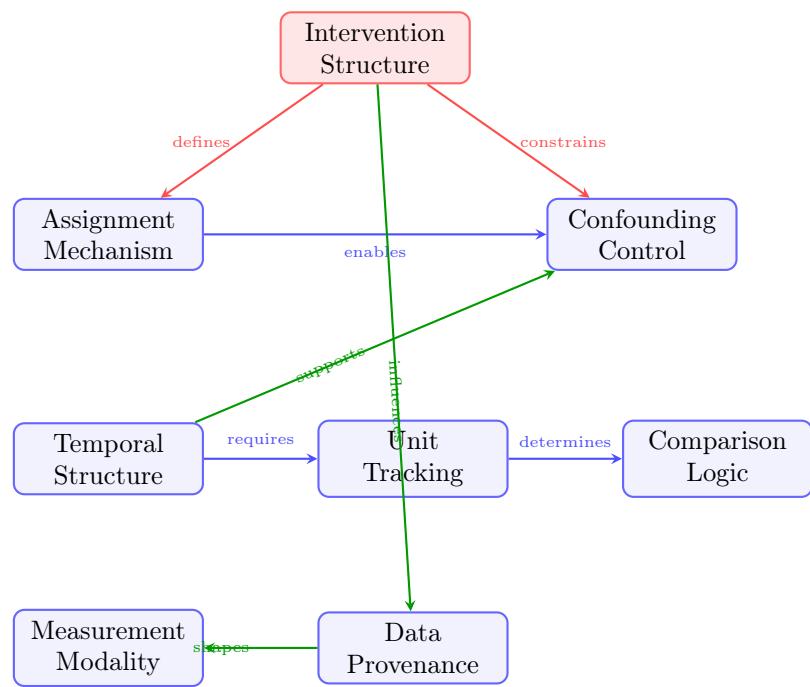


Figure 1: Taxonomy of Research Design Dimensions Organized by Conceptual Function

6 Design Selection and Justification

Selecting an appropriate research design involves mapping the research question to a feasible and inferentially defensible configuration of design dimensions. This process requires:

1. **Clarifying the inferential goal:** Is the objective causal identification, descriptive characterization, prediction, or hypothesis generation? What population and effect are of interest?
2. **Identifying constraints:** What ethical, practical, temporal, or resource constraints limit design options? Are randomization or experimental manipulation

**Dependency Types:**

- Logical (definitional)
- Practical (common)
- Inferential (analytical)

Figure 2: Key Interdependencies Among Design Dimensions

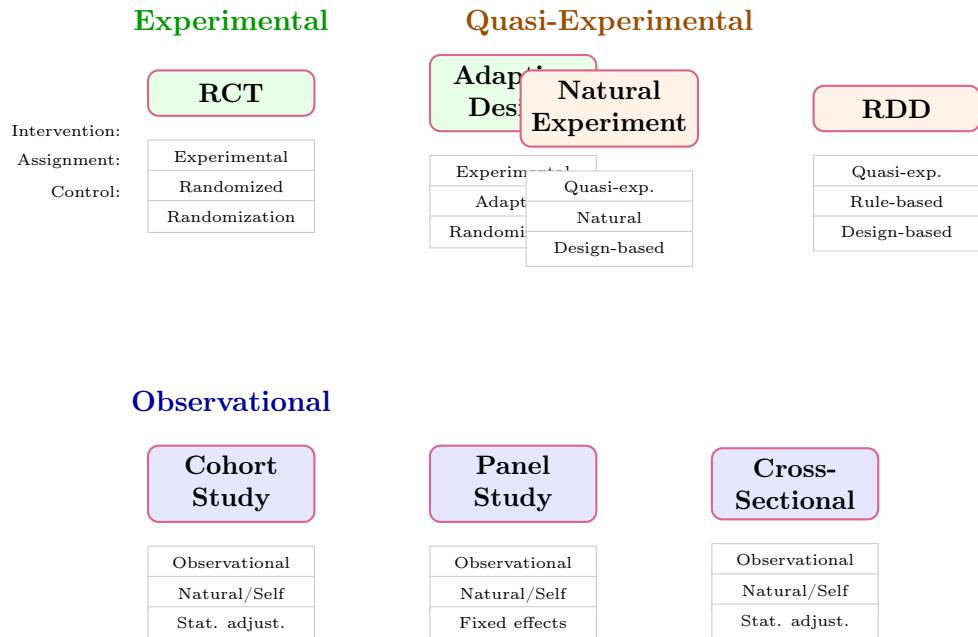


Figure 3: Major Research Schemes and Their Dimensional Configurations

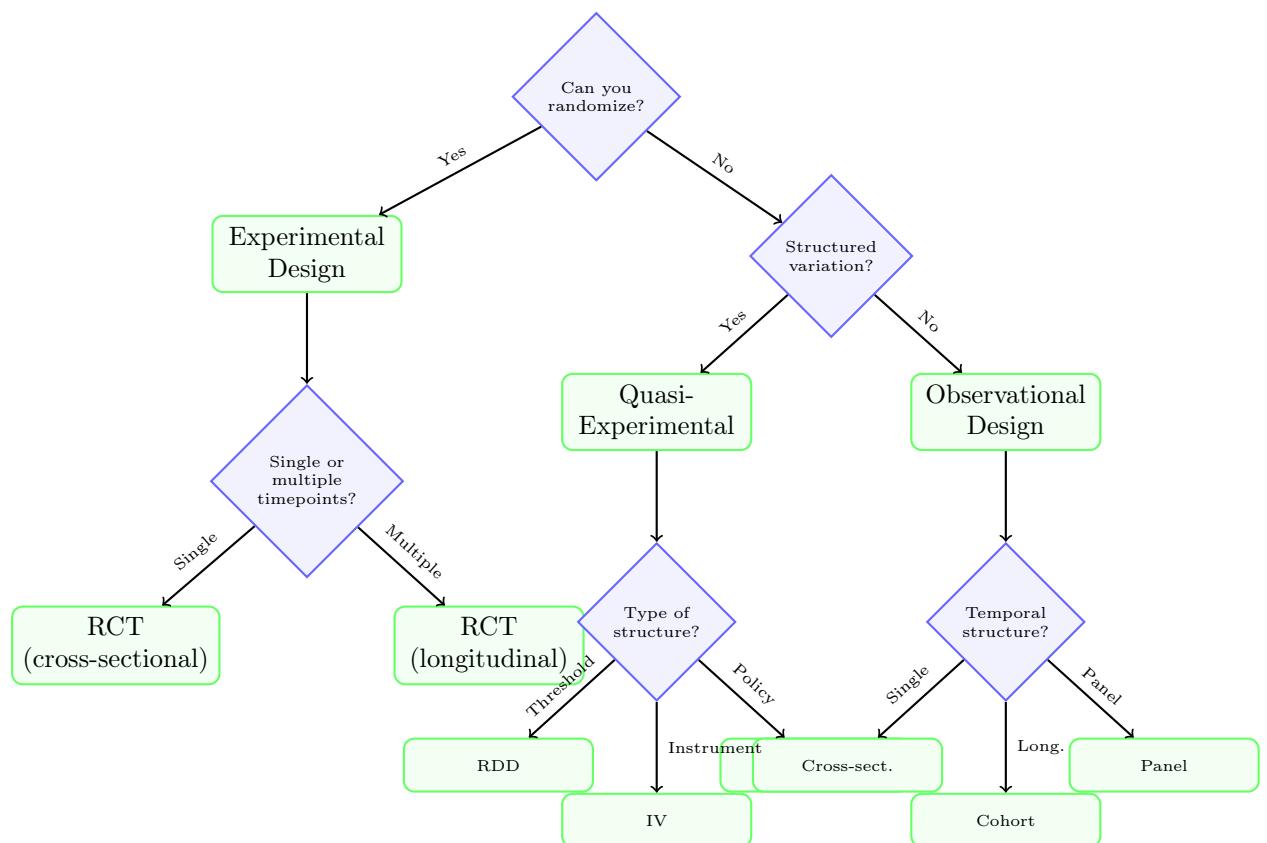


Figure 4: Design Space Navigation: How Key Dimensional Choices Lead to Canonical Research Schemes

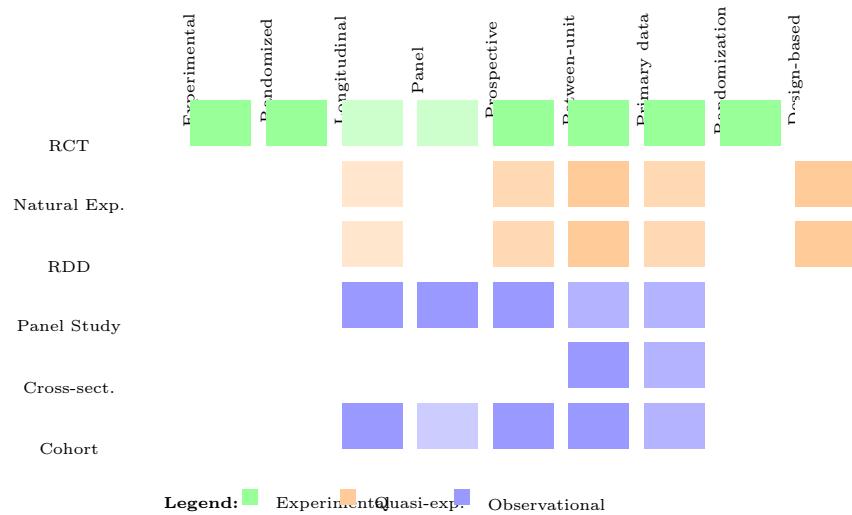


Figure 5: Scheme-Dimension Matrix: Characteristic Values for Major Research Schemes

From Dimensions to Inference

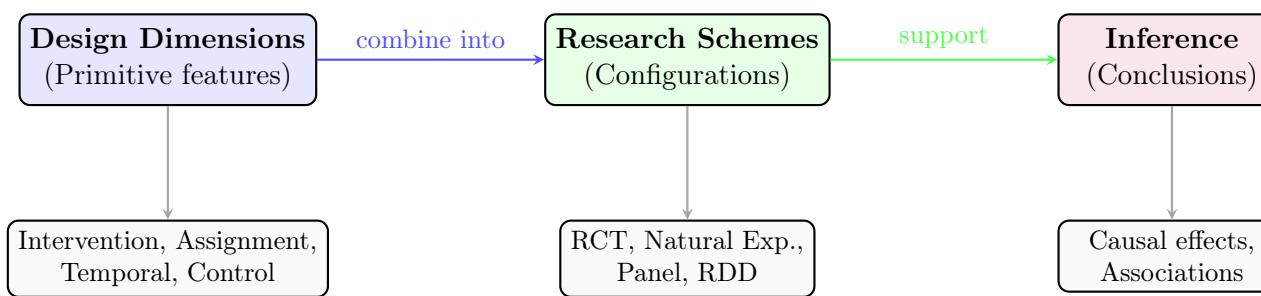


Figure 6: Conceptual Framework: How Design Dimensions Combine into Research Schemes to Support Inference

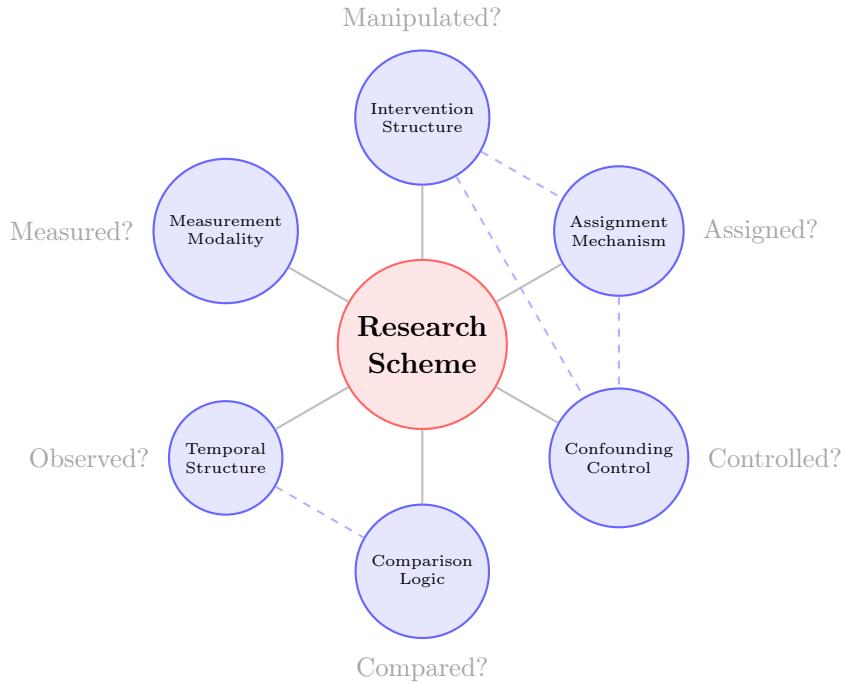


Figure 7: Core Design Dimensions as an Interconnected System

feasible? What measurement modalities are available and valid for key constructs?

3. **Assessing identification strategies:** What assumptions are required for valid inference under each feasible design? Which confounding control strategies are available given the data structure?
4. **Evaluating tradeoffs:** How do different designs balance internal validity, external validity, precision, bias, measurement validity, and ethical considerations?
5. **Justifying the chosen design:** Explicitly articulate why the selected configuration is appropriate given the research context, what assumptions it relies upon, and why alternative configurations are less suitable.

Design justification is not merely procedural. It requires substantive argumentation about why the chosen dimensional configuration supports the intended inference and why alternative configurations are less suitable. This justification should address both the strengths and limitations of the design, including the assumptions on which valid inference depends.

6.1 Design Justification Template

A complete design justification should address:

- **Research question and inferential goal:** What is the target estimand? Is causal identification required?
- **Intervention structure justification:** Why is the chosen level of intervention (experimental, quasi-experimental, observational) appropriate or necessary?

- **Assignment mechanism rationale:** How does the assignment mechanism support or constrain causal inference?
- **Temporal design rationale:** Why is the chosen temporal structure sufficient for the research question?
- **Measurement validity argument:** Why are the chosen measurement modalities valid for the constructs of interest? What are their known biases and how are they addressed?
- **Confounding control justification:** What confounding control strategy is employed? What assumptions does it require? Why are these assumptions plausible in this context? What sensitivity analyses or robustness checks are conducted?
- **Alternative designs considered:** What other dimensional configurations were considered? Why were they rejected (infeasibility, ethical concerns, weaker identification)?
- **Limitations acknowledged:** What are the key limitations of the chosen design? What threats to validity remain? How might these limitations affect interpretation?

Transparent articulation of design logic strengthens the credibility of empirical claims and enables critical evaluation by the scientific community.

6.2 Example: Design Justification for a Cohort Study of Educational Interventions

Research question: Does participation in an afterschool mentoring program improve high school graduation rates among at-risk youth?

Dimensional configuration: Observational design with self-selection into mentoring; longitudinal temporal structure following students from program entry through graduation-eligible age; cohort tracking of students who entered high school in the same year; prospective directionality with program participation measured at baseline and graduation observed subsequently; between-unit comparison of program participants and non-participants; primary data collection through surveys and administrative records; statistical adjustment using propensity score weighting to control confounding.

Justification:

- Random assignment to mentoring is infeasible because the program operates outside researcher control and denying mentoring to willing participants raises ethical concerns.
- Longitudinal temporal structure is necessary to establish temporal precedence and observe graduation outcomes, which occur years after program entry.
- Cohort design ensures comparability in terms of entry year, reducing confounding by cohort-specific factors (e.g., policy changes, economic conditions).
- Prospective directionality minimizes recall bias in program participation measurement.

- Propensity score weighting controls for observed confounders (prior academic performance, socioeconomic status, family structure, neighborhood characteristics) that predict both program participation and graduation. Conditional independence is plausible if all key confounders are measured.
- Administrative records for graduation outcomes reduce measurement error relative to self-report.
- Limitations: Unmeasured confounding remains possible (e.g., student motivation, family support). Sensitivity analyses quantify the strength of unmeasured confounding needed to eliminate the observed effect. External validity is limited to the specific program and student population studied.

This justification explicates the logic linking design choices to inferential goals, acknowledges assumptions and limitations, and demonstrates considered decision-making.

7 Recap: From Research Questions to Research Schemes

The relationship between research questions and research schemes is not mechanical. Researchers do not simply "apply" a predetermined scheme to a given question. Instead, design selection emerges from an iterative process of conceptualization, operationalization, and constraint mapping. This section describes the practical workflow through which researchers move from substantive questions to feasible research designs.

7.1 The Design Selection Workflow

Research design begins with a substantive question and proceeds through several stages of refinement and constraint assessment. The process is rarely linear—researchers often cycle between stages as they discover new constraints or refine their understanding of what is feasible and inferentially defensible.

7.1.1 Stage 1: Conceptual Clarification

Before selecting a design, researchers must clarify what they are trying to learn. This involves:

Articulating the substantive question. What real-world phenomenon or relationship are you trying to understand? For example: "Do afterschool programs improve academic outcomes?" or "How does social media use affect mental health?"

Identifying the type of inference desired. Different research goals require different designs:

- *Causal inference:* Estimating the effect of an intervention or exposure on an outcome (e.g., "Does X cause Y?")
- *Descriptive inference:* Characterizing a population, distribution, or relationship (e.g., "What proportion of X exhibit Y?")

- *Predictive inference:* Forecasting future outcomes or states (e.g., "Given X, what is the likely value of Y?")
- *Exploratory inquiry:* Generating hypotheses or identifying patterns (e.g., "What factors are associated with Y?")

The inferential goal fundamentally constrains design options. Causal inference requires temporal precedence and confounding control. Descriptive inference may not require these features but demands representative sampling. Prediction prioritizes out-of-sample performance over causal interpretation.

Defining the target population and estimand. For whom or what should the inference hold? The target population determines sampling requirements. The estimand (the quantity to be estimated) determines what comparisons must be constructed and what variation must be observed.

7.1.2 Stage 2: Operationalization

Conceptual clarity must be translated into measurable variables and observable units. This stage involves:

Defining units of observation and analysis. What are the entities being studied? Individuals? Organizations? Geographic areas? Time periods? The unit definition affects what data sources are available and what level of aggregation is appropriate.

Operationalizing key constructs. How will abstract concepts (e.g., "program participation," "academic achievement," "mental health") be measured? Available measurement modalities constrain design options. If valid biomarkers exist, experimental designs with objective measurement become more feasible. If only self-report is available, designs must account for reporting bias.

Identifying the exposure or treatment of interest. What is the intervention, exposure, or predictor whose effect or association is of interest? Can it be manipulated? Is it binary or continuous? Does it vary naturally? These characteristics determine whether experimental, quasi-experimental, or observational designs are possible.

Specifying the outcome. What is being affected or predicted? When can it be observed? How long after exposure does it occur? The temporal gap between exposure and outcome determines the required duration of follow-up and whether prospective or retrospective observation is feasible.

At this stage, researchers often discover that their initial question is not operationalizable with available measures. This may require reconceptualizing the question, seeking new measurement tools, or accepting that some aspects of the question cannot be empirically addressed.

7.1.3 Stage 3: Constraint Assessment

With a clear question and operational definitions, researchers assess what designs are feasible given practical, ethical, and resource constraints:

Ethical constraints. Can the exposure be manipulated? Is random assignment ethically acceptable? Harmful exposures cannot be experimentally assigned. Withholding beneficial interventions from control groups may be unethical. These constraints rule out certain experimental designs.

Feasibility constraints. Can the researcher control assignment? Do the necessary data exist? Can subjects be followed over time? Many exposures (race, gender, early-life experiences) cannot be manipulated. Many outcomes (rare diseases, long-term effects) cannot be observed within typical research timelines. These constraints often necessitate observational designs.

Resource constraints. What is affordable in terms of time, funding, and personnel? Experimental designs with primary data collection are expensive. Longitudinal designs require sustained funding. Large representative samples are costly. Resource constraints may force compromises between ideal and feasible designs.

Temporal constraints. How quickly are results needed? Experimental trials require time for recruitment, intervention, and follow-up. Longitudinal designs span years or decades. Secondary data analysis can be rapid but sacrifices control over measurement. Urgent questions may necessitate cross-sectional designs despite inferential limitations.

Data availability. What data already exist? Administrative records may enable large-scale quasi-experimental designs at low cost but with construct validity concerns. Survey data may provide rich measurement but limited sample sizes. The existence and accessibility of secondary data fundamentally shapes design options.

Constraint assessment often reveals that the ideal design for the research question is infeasible. This forces researchers to select among imperfect alternatives, balancing inferential strength against practical realities.

7.1.4 Stage 4: Dimensional Configuration

Given the clarified question and assessed constraints, researchers now configure design dimensions to construct a feasible scheme:

Intervention structure. If manipulation is ethical and feasible, experimental designs are possible. If manipulation is impossible but structured natural variation exists (policy changes, thresholds, instruments), quasi-experimental designs may be viable. Otherwise, observational designs are necessary.

Assignment mechanism. If intervention is possible, can randomization be implemented? If not, what determines assignment? The assignment mechanism determines what identification strategies are available and what assumptions must be invoked.

Temporal structure and unit tracking. Can units be followed over time? Is repeated observation feasible given attrition and cost? The temporal dimension is determined by the need to establish temporal precedence, observe change, or exploit within-unit variation for confounding control.

Comparison logic. What source of counterfactual comparison is most credible? Between-unit comparisons require exchangeability (achieved through randomization, matching, or adjustment). Within-unit comparisons require temporal stability. The comparison logic follows from the assignment mechanism and available data.

Confounding control strategy. Given the assignment mechanism, temporal structure, and available covariates, what confounding control is possible? Randomization eliminates confounding by design. Design-based identification (RDD, IV, DiD) exploits structural features. Statistical adjustment relies on measured confounders. The control strategy is constrained by prior dimensional choices.

Measurement modality and data provenance. What measurement approaches are valid and feasible for key constructs? Primary data allow tailored measurement but at high cost. Secondary data are cheaper but may not align well with constructs. The measurement dimension is constrained by budget, construct definitions, and data availability.

This configuration process is iterative: choices on one dimension constrain options on others. For example, choosing longitudinal temporal structure enables fixed effects confounding control, which in turn may reduce the need for rich covariate measurement. Researchers navigate this constrained design space, seeking coherent configurations that support the intended inference.

7.1.5 Stage 5: Scheme Recognition and Adaptation

Once dimensional values are configured, researchers often recognize that their design corresponds to a canonical research scheme. This recognition is valuable because schemes come with established inferential frameworks, known assumptions, and standard methods.

For example, a design with:

- Observational intervention
- Self-selection assignment
- Longitudinal temporal structure
- Cohort tracking
- Prospective directionality
- Between-unit comparison
- Primary data
- Statistical adjustment for confounding

is recognizable as a *cohort study*. This recognition connects the design to a literature on cohort study methods, standard analysis approaches, and typical threats to validity.

However, many real designs do not perfectly match canonical schemes. Researchers may combine features from multiple schemes or create hybrid designs tailored to specific contexts. The dimensional framework enables precise communication about such designs even when they do not fit standard categories.

7.2 Justifying Design Choices

Design justification articulates the logic connecting research question, constraints, dimensional configuration, and inferential assumptions. A complete justification addresses:

Why this inferential goal? What substantive question motivates the study? What type of inference (causal, descriptive, predictive) is required to address it?

Why these operational definitions? How do the chosen measures capture the constructs of interest? What measurement biases exist and how are they addressed or acknowledged?

Why this dimensional configuration? For each dimension, explain:

- What alternatives were considered
- What constraints ruled them out or made them less desirable
- How the chosen value supports the inferential goal

What assumptions does this design require? Make explicit the identifying assumptions on which valid inference depends. For example:

- Experimental designs assume compliance and no spillover
- Quasi-experimental designs have design-specific assumptions (parallel trends for DiD, continuity for RDD, exclusion restriction for IV)
- Observational designs with statistical adjustment assume conditional independence (no unmeasured confounding)

How plausible are these assumptions? Provide substantive arguments or empirical evidence supporting key assumptions. Discuss what would have to be true for assumptions to be violated and whether such violations seem likely.

What are the key limitations? Acknowledge what the design cannot do. What threats to validity remain unaddressed? How might violations of assumptions or other limitations affect interpretation?

What sensitivity analyses are planned? How will robustness of findings to assumption violations be assessed? For observational designs, how strong would unmeasured confounding need to be to overturn results?

7.3 Example: From Question to Design

To illustrate the workflow, consider the development of a study design for the question: "Does participation in afterschool mentoring programs improve high school graduation rates among at-risk youth?"

Stage 1: Conceptual clarification.

- Inferential goal: Causal inference (effect of mentoring on graduation)
- Target population: At-risk youth (operationally defined as students with low prior achievement or socioeconomic disadvantage)
- Estimand: Average treatment effect of mentoring participation on graduation probability

Stage 2: Operationalization.

- Units: Individual students
- Exposure: Participation in specific mentoring program (binary: participant vs. non-participant)
- Outcome: High school graduation (binary: graduate vs. not), observed 4-5 years after program entry
- Key covariates: Prior academic performance, socioeconomic status, family structure, neighborhood characteristics (available from school records and baseline survey)

Stage 3: Constraint assessment.

- Ethical constraints: Random assignment infeasible—program operates independently and denying mentoring to willing students is ethically problematic
- Feasibility: Cannot manipulate program participation; program enrollment is voluntary
- Resources: Moderate budget allows baseline survey and administrative data linkage but not intensive primary data collection
- Temporal: Results needed within 6 years; graduation outcomes observable within this window
- Data: Administrative graduation records available; baseline survey feasible; no biomarkers needed

Stage 4: Dimensional configuration.

- Intervention: Observational (cannot randomize)
- Assignment: Self-selection (students choose whether to participate)
- Temporal: Longitudinal (must follow students from program entry to graduation)
- Unit tracking: Cohort (follow students entering high school in same year to control cohort effects)
- Directionality: Prospective (measure participation at baseline, observe graduation later)

- Comparison: Between-unit (compare participants to non-participants)
- Data provenance: Mixed (baseline survey + administrative graduation records)
- Measurement: Self-report for participation, administrative records for graduation
- Confounding control: Statistical adjustment via propensity score weighting (observed confounders predict both participation and graduation)

Stage 5: Scheme recognition. This configuration is recognizable as a prospective cohort study with propensity score adjustment. The design follows established cohort study methods while adapting to the specific context (mentoring programs, educational outcomes).

Justification.

- Random assignment ruled out by ethical and feasibility constraints
- Longitudinal structure necessary to observe graduation outcomes and establish temporal precedence
- Cohort design controls for cohort-specific confounders (policy changes, economic conditions)
- Propensity score weighting addresses selection bias by balancing groups on observed characteristics
- Key assumption: Conditional independence—all important confounders measured (plausible given rich administrative data and baseline survey)
- Limitations: Unmeasured confounding possible (e.g., student motivation, family support not fully captured). Sensitivity analyses will quantify robustness to unmeasured confounding. External validity limited to this program and population.

This example shows how design selection proceeds from question to constraints to dimensional configuration to scheme recognition, with justification explaining the logic at each step.

7.4 Common Design Selection Patterns

Certain patterns recur in design selection workflows. Recognizing these patterns helps researchers navigate the design space more efficiently.

Pattern 1: Ideal design infeasible → closest feasible alternative. Researchers often begin with an ideal design (typically an RCT for causal questions) but discover it is infeasible. The task becomes finding the closest approximation that remains feasible. This might mean:

- Experimental → Quasi-experimental (exploiting natural experiments, RDD, or IV when randomization is impossible)

- Randomized → Matching/adjustment (when assignment cannot be controlled but rich covariates are available)
- Longitudinal → Cross-sectional (when follow-up is impossible, accepting limitations on temporal inference)

Pattern 2: Constraint-driven dimension fixing. Some constraints directly determine dimensional values, which then cascade to constrain other dimensions:

- Cannot manipulate exposure → Observational intervention → Self-selection or natural assignment → Need confounding control through matching/adjustment or design-based identification
- Cannot follow units over time → Cross-sectional temporal structure → Cannot use within-unit comparison or fixed effects → Must rely on between-unit comparisons
- Only secondary data available → Predetermined measurement modality and data provenance → Construct validity concerns → May need validation studies

Pattern 3: Opportunistic design. Sometimes researchers discover unexpected opportunities that enable stronger designs:

- Policy change creates natural experiment enabling DiD or synthetic control
- Threshold-based program eligibility enables RDD
- Lottery-based oversubscribed program enables IV design
- Existing high-quality administrative data enables large-scale observational analysis

Opportunistic designs require recognizing these structural features and adapting methods to exploit them.

Pattern 4: Hybrid and adaptive designs. Researchers increasingly combine features from multiple canonical schemes:

- RDD with difference-in-differences (exploiting both threshold and panel structure)
- Matching combined with regression adjustment (doubly robust estimation)
- Natural experiment with synthetic controls (for improved comparison group construction)
- Experimental design with administrative data linkage (reducing measurement cost and attrition)

These patterns reflect the reality that design selection is a pragmatic problem-solving process, not a formulaic application of standard templates.

7.5 Moving Forward: Design to Analysis

Once a research scheme is selected and justified, researchers proceed to:

- Specifying the statistical model or estimation strategy appropriate for the scheme
- Determining sample size and power requirements
- Developing data collection protocols (for primary data) or data access strategies (for secondary data)
- Planning sensitivity analyses and robustness checks
- Anticipating and planning for threats to validity specific to the chosen design

These implementation details build on the dimensional configuration established during design selection. The dimensional framework provides the conceptual foundation; subsequent chapters on estimation, inference, and validity build the technical superstructure.

The key insight is that research design is not a matter of selecting a pre-packaged scheme from a menu. It is a process of navigating a constrained design space, configuring dimensions to balance inferential goals against practical realities, and justifying the resulting configuration through transparent articulation of assumptions and limitations.

8 Case Studies: Research Schemes in Practice

This section presents real-world examples of canonical research schemes, illustrating how dimensional configurations are realized in actual studies. Each case study identifies the research question, dimensional configuration, key design features, and main findings, demonstrating how abstract design principles translate into concrete empirical work.

8.1 Randomized Controlled Trials

8.1.1 Case Study 1: Moving to Opportunity Experiment

Research Question. Does moving from high-poverty to low-poverty neighborhoods improve economic and health outcomes for families?

Background. The Moving to Opportunity (MTO) experiment, conducted by the U.S. Department of Housing and Urban Development from 1994-1998, examined whether residential mobility could improve outcomes for families living in public housing in high-poverty neighborhoods.

Dimensional Configuration.

- **Intervention:** Experimental (researchers controlled housing voucher assignment)
- **Assignment:** Randomized (lottery-based assignment to three groups)
- **Temporal:** Longitudinal (followed families for 10-15 years)
- **Unit tracking:** Panel (same families tracked over time)

- **Directionality:** Prospective (voucher assignment preceded outcome measurement)
- **Comparison:** Between-unit (experimental vs. control groups)
- **Data provenance:** Mixed (baseline surveys, administrative records, follow-up surveys)
- **Measurement:** Administrative records (earnings, education), surveys (health, well-being), direct assessment (children's achievement tests)
- **Confounding control:** Randomization-based

Key Design Features.

- 4,604 low-income families with children randomized to: (1) experimental voucher (must move to low-poverty area), (2) Section 8 voucher (no location restriction), or (3) control (no voucher)
- Compliance issues: Not all families assigned vouchers used them; not all experimental group families moved to low-poverty areas
- Long-term follow-up enabled assessment of effects on children who were young at randomization

Key Findings. Moving to lower-poverty neighborhoods produced mixed results: improved mental health and subjective well-being for adults, reduced obesity and diabetes, but little effect on economic self-sufficiency. For children, effects depended critically on age at move: children who moved before age 13 had substantially higher earnings as adults, while those who moved as adolescents showed no earnings gains.

Lessons. This case illustrates: (1) the power of randomization to answer causal questions that would be confounded in observational studies (families who voluntarily move differ from those who don't), (2) the importance of long-term follow-up for outcomes that take years to manifest, (3) compliance challenges even in experimental designs, and (4) the value of examining heterogeneous treatment effects by subgroup (age at move).

8.1.2 Case Study 2: Oregon Health Insurance Experiment

Research Question. What is the causal effect of health insurance coverage on health outcomes and healthcare utilization?

Background. In 2008, Oregon expanded its Medicaid program through a lottery due to budget constraints, creating a rare opportunity for a randomized evaluation of health insurance.

Dimensional Configuration.

- **Intervention:** Experimental (lottery-based selection for Medicaid)
- **Assignment:** Randomized (lottery determined eligibility)
- **Temporal:** Longitudinal (baseline and follow-up measurements)
- **Comparison:** Between-unit (lottery winners vs. losers)
- **Data provenance:** Mixed (administrative hospital records, credit reports, in-person health assessments, surveys)
- **Measurement:** Biomarkers (blood pressure, cholesterol, blood sugar), administrative records (hospital visits, credit), self-report (health status, access)
- **Confounding control:** Randomization via lottery

Key Findings. Medicaid coverage increased healthcare utilization, reduced financial strain (fewer medical debts), and improved self-reported health and mental health. However, no statistically significant effects were detected on measured physical health outcomes (blood pressure, cholesterol, glycated hemoglobin) after two years, though confidence intervals included clinically meaningful effects.

Lessons. Natural policy variation can create experimental opportunities. Biomarker measurement provided objective health assessment beyond self-report. The null findings on some outcomes were scientifically valuable, demonstrating that causal effects may differ from observational associations.

8.2 Quasi-Experimental Designs

8.2.1 Case Study 3: Minimum Wage and Employment (Card & Krueger)

Research Question. Does raising the minimum wage reduce employment, as predicted by standard economic theory?

Background. In 1992, New Jersey raised its minimum wage from \$4.25 to \$5.05 while neighboring Pennsylvania's minimum wage remained unchanged, creating a natural experiment.

Dimensional Configuration.

- **Intervention:** Quasi-experimental (policy change not controlled by researchers)
- **Assignment:** Natural (state-level policy determined exposure)
- **Temporal:** Longitudinal (before and after wage increase)
- **Comparison:** Between-unit (NJ vs. PA) and within-unit (same restaurants before/after)
- **Data provenance:** Primary (telephone surveys of fast-food restaurants)
- **Confounding control:** Design-based (difference-in-differences)

Key Design Features.

- Difference-in-differences design: compared change in employment in NJ (treatment) to change in PA (control)
- Parallel trends assumption: NJ and PA would have followed similar employment trends absent the wage increase
- Focus on fast-food industry where minimum wage workers are concentrated

Key Findings. Contrary to standard predictions, employment in NJ fast-food restaurants increased slightly relative to PA following the wage increase. No evidence that minimum wage increase reduced employment.

Lessons. Natural policy variation can enable causal inference when randomization is impossible. DiD design controls for state-specific time-invariant factors and common time trends. Geographic comparison units (neighboring states) increase plausibility of parallel trends. Findings challenged conventional wisdom and sparked extensive methodological and substantive debate.

8.2.2 Case Study 4: Class Size and Achievement (Angrist & Lavy, RDD)

Research Question. Does reducing class size improve student achievement?

Background. Israeli schools follow a rule limiting class size to 40 students (Maimonides' rule). When enrollment exceeds 40, schools must create two classes, creating sharp discontinuities in class size at enrollment thresholds (40, 80, 120, etc.).

Dimensional Configuration.

- **Intervention:** Quasi-experimental (rule-based class assignment)
- **Assignment:** Rule-based (deterministic function of enrollment)
- **Temporal:** Cross-sectional (single year of data)
- **Comparison:** Between-unit (students just above vs. just below thresholds)
- **Data provenance:** Secondary (administrative school records)
- **Measurement:** Administrative test scores
- **Confounding control:** Design-based (regression discontinuity)

Key Design Features.

- Discontinuity design: students in schools just above threshold (e.g., 41 students → two classes of 20) compared to just below (e.g., 40 students → one class of 40)
- Local comparison: estimates class size effect for students near enrollment thresholds
- Continuity assumption: student characteristics smooth across threshold

Key Findings. Reducing class size from 40 to 20-25 students improved test scores, particularly in reading. Effects larger for disadvantaged students.

Lessons. Institutional rules create quasi-random variation exploitable for causal inference. RDD provides highly credible local causal estimates but limited generalizability beyond threshold region. Administrative data enable large-scale analysis at low cost.

8.2.3 Case Study 5: Military Service and Earnings (Angrist, IV Design)

Research Question. What is the effect of military service on subsequent civilian earnings?

Background. During the Vietnam War, draft eligibility was determined by a lottery based on birth date, creating exogenous variation in military service.

Dimensional Configuration.

- **Intervention:** Quasi-experimental (draft lottery)
- **Assignment:** Randomized instrument (lottery) but not randomized treatment (service)
- **Temporal:** Longitudinal (lottery in 1970s, earnings observed in 1980s)
- **Comparison:** Between-unit (high vs. low lottery numbers)
- **Data provenance:** Secondary (Social Security earnings records)
- **Confounding control:** Design-based (instrumental variables)

Key Design Features.

- Draft lottery as instrument: randomly assigned lottery number strongly predicts military service but should not directly affect earnings
- Estimates local average treatment effect (LATE) for compliers: men induced to serve by the draft
- Does not estimate effect for volunteers or those who would avoid service regardless of draft

Key Findings. Military service reduced subsequent civilian earnings by about 15% for white veterans. Effects concentrated among those who served in combat and those who would have attended college absent the draft.

Lessons. IV designs exploit random or as-if random instruments to address endogeneity. Estimates apply to specific subpopulation (compliers). Natural experiments from policy variation can answer causal questions decades later when combined with administrative data.

8.3 Observational Designs

8.3.1 Case Study 6: Framingham Heart Study (Cohort)

Research Question. What factors contribute to cardiovascular disease?

Background. Launched in 1948, the Framingham Heart Study recruited 5,209 adults in Framingham, Massachusetts for long-term follow-up to identify risk factors for heart disease.

Dimensional Configuration.

- **Intervention:** Observational
- **Assignment:** Self-selection (voluntary participation)
- **Temporal:** Longitudinal (decades of follow-up, now multi-generational)
- **Unit tracking:** Cohort (same individuals followed repeatedly)
- **Directionality:** Prospective (risk factors measured before disease onset)
- **Comparison:** Between-unit (those who develop disease vs. those who don't)
- **Data provenance:** Primary (repeated medical examinations every 2 years)
- **Measurement:** Biomarkers (blood pressure, cholesterol), direct observation (physical exams), medical records (disease outcomes)
- **Confounding control:** Statistical adjustment for multiple risk factors

Key Findings. Identified major cardiovascular risk factors including high blood pressure, high cholesterol, smoking, obesity, diabetes, and physical inactivity. Established the concept of "risk factors" in epidemiology.

Lessons. Prospective cohort designs establish temporal precedence and reduce recall bias. Long-term follow-up enables study of diseases with long latency periods. Cannot establish causality with certainty (residual confounding possible) but provides strong evidence when combined with biological plausibility and consistency across studies. Multi-generational design allows study of genetic and familial factors.

8.3.2 Case Study 7: Nurses' Health Study (Cohort)

Research Question. What lifestyle factors affect women's health, particularly cancer and cardiovascular disease?

Background. Initiated in 1976, enrolled 121,700 female nurses aged 30-55. Nurses chosen because expected to provide accurate health information and high follow-up rates.

Dimensional Configuration.

- **Intervention:** Observational
- **Assignment:** Self-selection into exposures (diet, hormone use, lifestyle)
- **Temporal:** Longitudinal (ongoing since 1976)
- **Unit tracking:** Cohort (same nurses followed via biennial questionnaires)
- **Directionality:** Prospective
- **Data provenance:** Primary (questionnaires) plus medical records for outcome validation
- **Measurement:** Self-report (diet, lifestyle) validated by biomarkers in subsample
- **Confounding control:** Statistical adjustment, stratification

Key Findings. Documented health effects of oral contraceptives, hormone replacement therapy, diet, physical activity, and smoking. Findings influenced clinical practice and public health recommendations.

Lessons. Selective sampling (nurses) traded representativeness for measurement quality and retention. Very large sample size enabled detection of small effects and analysis of rare outcomes. Repeated measures allowed time-varying exposure assessment. Validation studies addressed measurement error in self-reported exposures.

8.3.3 Case Study 8: Smoking and Lung Cancer (Case-Control)

Research Question. Is smoking a cause of lung cancer?

Background. Early case-control studies in the 1950s (Doll & Hill, Wynder & Graham) compared smoking histories of lung cancer patients to controls.

Dimensional Configuration.

- **Intervention:** Observational
- **Assignment:** Outcome-based sampling (select cases and controls)
- **Temporal:** Cross-sectional with retrospective exposure assessment
- **Directionality:** Retrospective (outcome known, assess past exposure)
- **Comparison:** Between-unit (cases vs. controls)
- **Data provenance:** Primary (interviews)
- **Measurement:** Self-report of smoking history
- **Confounding control:** Matching on age, sex, hospital

Key Design Features.

- Efficient for rare outcomes (lung cancer uncommon in general population)
- Cases: patients with confirmed lung cancer
- Controls: patients hospitalized for other conditions (not lung disease)
- Matching ensured comparability on key demographics

Key Findings. Overwhelming association between smoking and lung cancer: nearly all lung cancer patients were smokers; dose-response relationship (more smoking → higher risk).

Lessons. Case-control design efficient for rare diseases. Retrospective exposure assessment vulnerable to recall bias, but smoking history relatively objective. Findings consistent across multiple independent studies strengthened causal inference despite observational design. Established foundation for subsequent prospective cohort studies and eventually experimental evidence from animal studies.

8.3.4 Case Study 9: Social Capital and Health (Cross-Sectional)

Research Question. Is social connectedness associated with health outcomes at the community level?

Background. Ecological studies examined whether communities with higher social capital (civic engagement, trust, social networks) have better health outcomes.

Dimensional Configuration.

- **Intervention:** Observational
- **Assignment:** Natural variation across communities
- **Temporal:** Cross-sectional (single time point)
- **Directionality:** Simultaneous (social capital and health measured concurrently)
- **Comparison:** Between-unit (across communities/states)
- **Unit of analysis:** Aggregate (states or communities)
- **Data provenance:** Secondary (surveys, vital statistics)
- **Measurement:** Survey-based social capital indices, administrative health data
- **Confounding control:** Statistical adjustment for socioeconomic factors

Key Findings. States with higher social capital showed lower mortality rates, even controlling for income and other socioeconomic factors. Associations found for multiple health outcomes.

Lessons. Cross-sectional ecological designs can identify community-level associations but cannot establish causality (reverse causation possible: health affects social capital). Cannot infer individual-level relationships (ecological fallacy). Useful for generating hypotheses and identifying targets for intervention but require confirmation from stronger designs.

8.4 Panel Studies

8.4.1 Case Study 10: Panel Study of Income Dynamics (PSID)

Research Question. How do economic circumstances change over time and across generations?

Background. Launched in 1968, PSID has followed U.S. families annually (now biennially), tracking economic mobility, family dynamics, and health across five decades.

Dimensional Configuration.

- **Intervention:** Observational
- **Assignment:** Self-selection into economic behaviors
- **Temporal:** Longitudinal (50+ years)
- **Unit tracking:** Panel (same families followed, including children who form new households)
- **Directionality:** Prospective
- **Comparison:** Within-unit and between-unit
- **Data provenance:** Primary (interviews)
- **Sampling:** Probability sample initially representative of U.S.
- **Measurement:** Self-report (income, employment, family structure)
- **Confounding control:** Fixed effects for within-unit comparisons

Key Applications. Researchers have used PSID to study: income volatility and mobility, effects of job loss and unemployment, intergenerational transmission of economic status, effects of family structure changes, and long-term effects of childhood poverty.

Key Findings. Documented substantial economic mobility within lifetimes but also persistent inequality across generations. Identified role of job displacement in long-term earnings losses. Showed childhood economic circumstances affect adult outcomes even controlling for education.

Lessons. Panel designs enable within-person analysis controlling for stable unobserved characteristics. Very long panel allows multi-generational analysis. Attrition challenges increase over time but can be addressed through weighting and bounds analysis. Panel data uniquely suited to questions about change, transitions, and dynamics.

8.5 Digital Trace Studies

8.5.1 Case Study 11: Social Media and Political Polarization

Research Question. Does exposure to opposing political views on social media reduce polarization?

Background. Researchers partnered with Twitter to conduct a field experiment where users were offered financial incentives to follow bot accounts that retweeted messages from elected officials and opinion leaders from the opposing political party.

Dimensional Configuration.

- **Intervention:** Experimental (researchers controlled bot follow recommendations)
- **Assignment:** Randomized (users randomly offered incentive or not)
- **Temporal:** Longitudinal (baseline and follow-up surveys, continuous Twitter activity)
- **Data provenance:** Mixed (Twitter behavioral data + surveys)
- **Measurement:** Digital trace (following behavior, tweeting, retweeting) + self-report (political attitudes)
- **Confounding control:** Randomization

Key Findings. Exposure to opposing views increased political polarization rather than reducing it, particularly among Republicans. Democrats showed no significant change.

Lessons. Digital platforms enable large-scale behavioral observation and field experiments. Combining trace data (actual behavior) with surveys (attitudes) provides richer picture than either alone. Demonstrates that intuitive predictions about social media effects may be wrong—empirical testing essential.

8.5.2 Case Study 12: Emotional Contagion on Facebook

Research Question. Do emotions spread through social networks?

Background. Facebook data scientists manipulated News Feed content for 689,003 users for one week, reducing exposure to either positive or negative emotional content.

Dimensional Configuration.

- **Intervention:** Experimental (Facebook controlled content shown)
- **Assignment:** Randomized (users assigned to conditions)
- **Temporal:** Short panel (one week)
- **Measurement:** Digital trace (linguistic analysis of posts)
- **Data provenance:** Secondary (platform data)
- **Confounding control:** Randomization

Key Findings. Reducing positive content led to more negative posts; reducing negative content led to more positive posts. Effect sizes were very small but statistically significant given large sample.

Lessons. Digital platforms enable massive-scale experiments. Automated text analysis allows outcome measurement without surveys. Study sparked major ethical debate about informed consent in platform experiments, illustrating that technical feasibility doesn't equal ethical acceptability. Very large samples can detect statistically significant effects that are substantively tiny.

8.6 Meta-Analysis

8.6.1 Case Study 13: Early Childhood Education Effects (Meta-Analysis)

Research Question. What are the overall effects of early childhood education programs on child development?

Background. Hundreds of studies have evaluated early childhood interventions with varying designs, populations, and findings. Meta-analysis synthesizes this evidence.

Dimensional Configuration.

- **Unit of analysis:** Studies (not individuals)
- **Data provenance:** Secondary (published and unpublished studies)
- **Temporal:** Retrospective synthesis
- **Measurement:** Effect sizes extracted from primary studies
- **Confounding control:** None directly; assessment of study quality

Key Design Features.

- Systematic literature search
- Inclusion criteria: experimental or quasi-experimental evaluations of early education
- Meta-regression to examine moderators (program type, child age, outcome domain, study design)
- Publication bias assessment

Key Findings. Early childhood programs show positive average effects on cognitive and social-emotional development. Effects larger for: higher-quality programs, disadvantaged children, and programs starting earlier. Effects persist into elementary school but may fade over time without sustained intervention.

Lessons. Meta-analysis provides quantitative synthesis across heterogeneous studies. Can examine moderators to understand when/where/for whom effects occur. Requires careful assessment of study quality and publication bias. Synthesis strength limited by quality of primary studies.

8.7 Synthesis: What Makes a Good Case Study?

These examples illustrate several principles:

Design follows question. The research question determines what type of design is needed. Causal questions typically require experimental or quasi-experimental designs; descriptive questions may use observational approaches.

Constraints shape design. Ethical and practical constraints often rule out ideal designs, forcing researchers to adapt. The best studies creatively exploit available opportunities (natural experiments, institutional rules, existing data).

Tradeoffs are inevitable. Every design sacrifices something: RCTs may lack external validity; quasi-experiments rely on strong assumptions; observational studies face confounding. Good research acknowledges these tradeoffs explicitly.

Context matters. Design appropriateness depends on the specific research context. Designs that work well for one question may be inappropriate for another.

Multiple studies build knowledge. Single studies rarely definitively answer questions. Convergent evidence from multiple designs (e.g., case-control, then cohort, then RCT for smoking and cancer) provides strongest causal inference.

Innovation in design is ongoing. Researchers continually develop new designs (synthetic controls, bunching designs, event studies) and combine features from multiple schemes to exploit specific opportunities.

Understanding these canonical examples provides a foundation for designing new studies and critically evaluating existing research.