

Contents

Acknowledgements	xi
I Introduction	1
1 Preamble	3
1.1 How to Read This Book	3
1.2 How to Work This Book	5
1.3 What This Book Will Not Do	5
2 What Is a Research Design?	6
2.1 MIDA: The Four Elements of a Research Design	6
2.2 Declaration, Diagnosis, Redesign	14
2.3 Example: A Decision Problem	17
2.4 Putting Designs to Use	22
3 Research Design Principles	24
4 Getting Started	28
4.1 Installing R	28
4.2 Declaration	29
4.3 Diagnosis	30
4.4 Redesign	31

4.5 Library of Designs 32
4.6 Long-Term Code Usability..... 32

II Declaration, Diagnosis, Redesign 33

5 Declaring Designs 35

5.1 Definition of Research Designs..... 35
5.2 Declaration in Code 38

6 Specifying the Model 41

6.1 Elements of Models 42
6.2 Types of Variables in Models 46
6.3 How to Specify Models..... 48
6.4 Summary..... 51

7 Defining the Inquiry 52

7.1 Elements of Inquiries 53
7.2 Types of Inquiries 58
7.3 How to Define Inquiries..... 61
7.4 Summary..... 64

8 Crafting a Data Strategy 65

8.1 Elements of Data Strategies 67
8.2 Challenges to Data Strategies 87
8.3 Summary..... 91

9 Choosing an Answer Strategy..... 92

9.1 Elements of Answer Strategies..... 92
9.2 Types of Answer Strategies..... 97
9.3 How to Choose an Answer Strategy 107
9.4 Summary..... 115

10	Diagnosing Designs	116
10.1	Elements of Diagnoses	118
10.2	Types of Diagnosands	122
10.3	Estimation of Diagnosands	124
10.4	How to Diagnose Designs.....	131
10.5	Summary.....	136
11	Redesigning	137
11.1	Redesigning over Data Strategies	137
11.2	Redesigning over Answer Strategies	143
11.3	Summary.....	146
12	Design Example	148
12.1	Declaration in Words.....	148
12.2	Declaration in Code	149
12.3	Diagnosis	151
12.4	Redesign	152
13	Designing in Code	154
13.1	Model	154
13.2	Inquiry	164
13.3	Data Strategy	166
13.4	Answer Strategy.....	169
13.5	Declaration	173
13.6	Diagnosis	175
13.7	Redesign	179
III	Research Design Library	181
14	Research Design Library	183

15	Observational : Descriptive	185
15.1	Simple Random Sampling	185
15.2	Cluster Random Sampling	190
15.3	Multilevel Regression and Poststratification	193
15.4	Index Creation	198
16	Observational : Causal	203
16.1	Process Tracing.....	203
16.2	Selection-on-Observables.....	208
16.3	Difference-in-Differences	211
16.4	Instrumental Variables	216
16.5	Regression Discontinuity Designs	221
17	Experimental : Descriptive	227
17.1	Audit Experiments.....	228
17.2	List Experiments	232
17.3	Conjoint Experiments	236
17.4	Behavioral Games.....	242
18	Experimental : Causal	249
18.1	Two-Arm Randomized Experiments	250
18.2	Block-Randomized Experiments.....	257
18.3	Cluster-Randomized Experiments	260
18.4	Subgroup Designs.....	263
18.5	Factorial Experiments	266
18.6	Encouragement Designs	271
18.7	Placebo-Controlled Experiments	279
18.8	Stepped-Wedge Experiments	283
18.9	Randomized Saturation Experiments	288
18.10	Experiments over Networks	292

19	Complex Designs	299
19.1	Discovery Using Causal Forests	299
19.2	Structural Estimation	305
19.3	Meta-analysis	310
19.4	Multi-site Studies	313
IV	Research Design Lifecycle	319
<hr/>		
20	Research Design Lifecycle	321
21	Planning	322
21.1	Ethics	322
21.2	Partners	326
21.3	Funding	329
21.4	Piloting	330
21.5	Criticism	333
21.6	Preanalysis Plan	334
22	Realization	338
22.1	Pivoting	338
22.2	Populated Preanalysis Plan	340
22.3	Reconciliation	341
22.4	Writing	344
23	Integration	347
23.1	Communicating	348
23.2	Archiving	349
23.3	Reanalysis	351
23.4	Replication	356
23.5	Meta-analysis	358

V Epilogue 361

24 Epilogue 363

VI References 365

Bibliography 367

Index 377

CHAPTER 1

Preamble

This book introduces a new way of thinking about research designs in the social sciences. Our hope is that this approach will make it easier to develop and to share strong research designs.

At the heart of our approach is the *MIDA* framework, in which a research design is characterized by four elements: a model, an inquiry, a data strategy, and an answer strategy. We have to understand each of the four on their own and also how they interrelate. The design encodes your beliefs about the world, it describes your questions, and it lays out how you go about answering those questions, in terms of both what data you collect and how you analyze it. In strong designs, choices made in the model and inquiry are reflected in the data and answer strategies, and vice versa.

We think of designs as objects that can be interrogated. Each of the four design elements can be “declared” in computer code and—if done right—the information provided is enough to “diagnose” the quality of the design through computer simulation. Researchers can then select the best design for their purposes by “redesigning” over alternative, feasible designs.

This way of thinking pays dividends at multiple points in the research design lifecycle: planning the design, implementing it, and integrating the results into the broader research literature. The declaration, diagnosis, and redesign process informs choices made from the beginning to the end of a research project.

1.1 How to Read This Book

We had multiple audiences in mind when writing this book. First, we were thinking of people looking for a high-level introduction to these ideas. If we only had 30 minutes with a person to try and communicate the core ideas, we would give them Part I. We were thinking of people who are new to the practice of research design and who are embarking on their first empirical projects. The *MIDA* framework introduced in Part I accommodates many different empirical approaches: qualitative and quantitative, descriptive and causal, observational and experimental. Beginners starting out in any of these traditions can use our framework to

consider how the design elements in those approaches fit together. We were also thinking of researchers-in-training: graduate students in seminar courses where the main purpose is to read papers and discuss the credibility of research findings. Such discussions can sometimes feel like a laundry list of complaints, but we hope our framework can focus attention on the most relevant issues. What, exactly, is the inquiry? Is it the right one to be posing? Are the data and answer strategies suited to the inquiry? We were also thinking of funders and decision-makers, who often wish to assess research in terms not of its results but of its design. Our approach provides a way of defining the design and diagnosing its quality.

Part II is more involved. We provide the formal foundations of the *MIDA* framework. We walk through each component of a research design in detail, describe the finer points of design diagnosis, and explain how to carry out a “redesign.” We hope Part II will resonate with several audiences of applied researchers both inside and outside of academia. We imagine it could be assigned early in a graduate course on research design in any of the social sciences. We hope data scientists and monitoring and evaluation professionals will find value in our framework for learning about research designs. Scholars will find value in declaring, diagnosing, and redesigning designs whether they are implementing randomized trials or multi-method archival studies, or calibrating structural theories with data.

In Part III, we apply the general framework to specific research designs. The result is a library of common designs. Many empirical research designs are included in the library, but not all. The set of entries covers a large portion of what we see in current empirical practice across social sciences, but it is not meant to be exhaustive.

We are thinking of three kinds of uses for entries in the design library. Collectively, the design entries serve to illustrate the fundamental principles of design. The entries clarify the variety of ways in which models, inquiries, data strategies, and answer strategies can be connected and show how high-level principles operate in common ways across very different designs. The second use is pedagogical. The library entries provide hands-on illustrations of designs in action. A researcher interested in understanding the “regression discontinuity design,” for example, can quickly see a complete implementation and learn under what conditions the standard design performs well or poorly. They can also compare the suitability of one type of design against another for a given problem. We emphasize that these descriptions of different designs provide entry points but they are not exhaustive, so we refer readers to recent methodological treatments of the different topics. The third use is as a starter kit to help readers get going on designs of their own. Each entry includes code for a basic design that can be fine-tuned to capture the specificities of particular research settings.

The last section of the book describes how our framework can help at different stages of the research process. Each of these sections should be readable for anyone who

has read Part I. The entry on preanalysis plans, for example, can be assigned in an experiments course as guidance for students filing their first preanalysis plan. The entry on research ethics could be discussed among coauthors at the start of a project. The entry on writing a research paper could be assigned to college seniors writing their first original research papers.

1.2 How to Work This Book

We will often describe research designs not just in words, but in computer code. If you want to work through the code and exercises, fantastic. This path requires investment in R, the `tidyverse`, and the `DeclareDesign` software package. Chapter 4 helps get you started. We think working through the code is very rewarding, but we understand that there is a learning curve. You could tackle the declaration, diagnosis, and redesign processes using any computer language you like,¹ but it is easier in `DeclareDesign` because the software guides you to articulate each of the four design elements.

If you want nothing to do with the code, you can skip it and just focus on the text. We have written the book so that understanding of the code is not required in order to understand research design concepts.

1.3 What This Book Will Not Do

This is a research design book, not a statistics textbook, or a cookbook with recipes applicable to all situations. We will not derive estimators, we will provide no guarantees of the general optimality of designs, and we will present no mathematical proofs. Nor will we provide all the answers to all the practical questions you might have about your design.

What we do offer is a language to express research designs. We can help you learn that language so you can describe your own design in it. When you can declare your design in this language, then you can diagnose it, figure out if it works the way you think it should, and then improve it through redesign.

¹On our Web site, we provide examples in R, Python, Stata, and Excel.

Index

A

Abadie, A., 111
Aguilar, R., 242
Ahram, A., 207
answer strategy, 12–13, 36, 48, 92–115; Bayesian, 102; characterization, 93; examples, 13; procedure, 96
archiving, 349
Arechar, A. A., 288
assignment: block randomized, 78–80; cluster randomized, 78–80; complete randomized, 78–79, 186; multi-arm random, 80; nonrandomized, 83–84; saturation, 78–80, 288–291; simple randomized, 80; stepped-wedge random, 82, 283–288; response-adaptative strategies, 83
attrition, 88–89
audit experiments, 228–232; design examples, 231–232
Avdeenko, A., 248
Average Treatment Effect (ATE), 36, 53, 108, 148, 267, 288
Average Treatment Effect on the Treated (ATT), 53, 58, 59, 214
Average Treatment Effect on the Untreated (ATU), 54, 58

B

Baron, H., 323
Bateson, R., 211
Bauer, P. C., 305
Baum, M. A., 305
Bayesian learning, 123
behavioral games, 242–248; design examples, 248
Berinsky, A. J., 266
bias, 123, 129
bias-eliminated coverage, 123
bias-variance trade-off, 143; illustration, 145
Birkelund, G., 231
Bisbee, J., 194
Blair, G., 312
Blair, R. A., 278

block-randomized experiments, 257–260; design examples, 260

Bonilla, T., 337
Borghard E., 208
Bradley, V. C., 189
Broockman, D., 202, 282

C

Carnegie, A., 226
Carreri, M., 216
Castillo, A. V., 278
causal forest designs, 299–305; design examples, 305
Christensen, D., 312
Clemm von Hohenberg, B., 305
cluster-randomized experiment, 260–263; design examples, 263
cluster random sample, 190–193; design examples, 193
Collins, J. E., 266
communicating, 348–349
compliance types, 87–88, 218, 272–273, 276–282
Complier Average Causal Effect (CACE), 54, 57, 273–278, 282
Conditional Average Treatment Effect (CATE), 57, 165, 251, 267–270, 343
conjoint experiments, 236–242; design examples, 241–242
Cook, T., 53
Cooper, J., 283
Coppock, A., 231, 236, 298, 310, 312
Coverage, 118, 123, 129, 131
criticism, 333–234
crossover design, 82–83
Cruz, C., 236

D

data strategy, 11–12, 36, 65–91; challenges, 87–91; elements, 67–85; examples, 12; illustration, 88
deception, 242, 247
de Chaisemartin and d'Haultfoeuille estimator, 212, 214–215

- declaration, diagnosis, redesign workflow, 14–17;
example, 17
- design declaration, 14, 35–40; in code, 29; complete,
118
- design diagnosis, 14–15, 116–136; analytic, 124–
126; by simulation, 126–130; under model
uncertainty, 133–135
- DesignLibrary, 32
- diagnosand, 15–16, 118; definition, 116, 118;
diagnosand-completeness, 48, 118; estimation,
124; examples, 123; statistics, 118–119
- diagnosis, 30
- diagnostic statistics, 15; of competing models, 135;
definition, 15, 118–119; examples, 122
- difference-in-differences, 211–216; design examples,
215–216
- Directed Acyclic Graphs (DAGs), 43–44
- discovery, 299–305; design examples, 305
- Druckman, J. N., 305
- Dube, O., 216
- Dunning, T., 317
- E**
- Ecker, U.K.H., 266
- Eckles, D., 288
- Egger, D., 292
- encouragement designs, 271–279; design examples,
278–279
- Epstein, Z., 288
- estimand, 10–11
- ethics, 322–329
- excludability assumption, 50, 89–90, 219, 274
- exogeneity assumption, 219, 274
- expected maximum cost, 123
- experiments over networks, 292–298; design
examples, 298
- F**
- factorial designs. *See* factorial experiments
- factorial experiments, 266–271; design examples,
270–271
- Fang, A., 232
- Farrer, B. D., 298
- Fisher, R. A., 111
- Flaxman, S., 189
- Foos, F., 113
- Francois, P., 310
- Frederiksen, K.V.S., 317
- Freedman, D., 255
- Frey, A., 310
- funding, 329
- G**
- Gelman, A., 116
- Gerber, A. S., 89, 113, 124–125, 285, 288, 327
- Gilligan, M., 248
- Green, D. P., 124–125, 263, 283, 298, 327
- Green, J., 305
- Gulzar, S., 148
- Guess, A., 232
- H**
- Haushofer, J., 292
- Humphreys, M., 76, 103, 232, 306, 323
- hypothesis tests, 99–102, 113–115, 120
- I**
- Imai, K., 125, 233, 259
- index creation, 198–202; design examples, 202
- instrumental variables, 216–221; design examples,
221
- inquiry, 10–11, 36, 52–64; causal, 53; causal attri-
bution, 58–60; complex counterfactual, 60;
with continuous causal variables, 60–61; data-
dependent, 58; descriptive, 36, 53, 55–56, 185;
examples, 11; outcomes, 9, 55; units, 9, 53–55;
values, 9, 54 (*see also* estimand)
- instrumental variables, 46, 203, 216–221; design
examples, 221
- Intention-to-treat (ITT), 218, 274–275
- interference, 87, 90–91
- interval estimators, 104–107
- Intra-cluster Correlation (ICC), 158, 190,
260–263
- Isakov, M., 189
- Iyengar, S., 248
- J**
- Jacobs, A., 76, 103, 204
- Jefferson, H., 202
- K**
- Kalla, J., 202, 259, 282
- Kao, K., 241
- Karpowitz, C. F., 270
- Khan, M., 148
- Kirkland, P., 226
- Krasno, J. S., 298
- Kuriwaki, S., 189
- L**
- Lancee, B., 231
- Larsen, E. N., 231
- Lax, J. R., 198
- Lazer, D., 305
- Lenoir, B., 298
- Lewandowsky, S., 266
- Lin, J., 305
- Lin, W., 255
- Lin estimator, 255–257
- list experiments, 232–236; design examples, 236
- Local Average Treatment Effect (LATE), 54, 222,
273
- López-Moctezuma, G., 310
- Lyll, J., 259, 323

M

matching. *See* selection-on-observables
maximum proportion of subjects harmed, 123, 323
measurement, 84–86; data-adaptative, 86–87;
multiple, 86; over-time, 86
measurement error, 85
Meng, X., 189
Merkley, E., 190
meta-analysis, 310–312; design examples, 312
MIDA: definition, 6–7; elements, 37; illustration, 7
Middleton, J. A., 125
Miguel, E., 292
Mills methods, 75–76
Minimum Detectable Effect (MDE), 15, 123,
134–135, 139–140
model, 41–42; causal models, 8, 35, 41–51; func-
tional relations, 35, 43–45; how to specify, 48–51;
inquiry models, 8, 65, 306, 307; in MIDA, 8,
35–36, 41–44, 48, 65, 306, 307; probability distri-
bution over exogenous variables, 45–46; reference
models, 8, 36, 41–44, 48, 65, 306, 307; as a set, 8;
signature, 35, 42–43; statistical models, 8, 113,
169; variables in, 46–47
monotonicity assumption, 50, 204, 207, 219–220, 274
Monte Carlo simulation, 116; for qualitative research,
76; in simulation error, 130
Montero, S., 310
Moscoso-Rojas, M., 278
Mosleh, M., 288
Mousa, S., 263, 344–345
Multi-level Regression and Poststratification (MRP),
193–198; design examples, 198
multi-site studies, 313–317; design examples, 317

N

Nellis, G., 221
Neyman standard errors, 122, 251
Niehaus, P., 292
noncompliance. *See* compliance types
noninterference assumption, 219, 274
nonresponse, 88, 188
nonzero effect assumption, 220, 274–275
null risk, 123

O

Ognyanova, K., 305

P

Paglayan, A., 215
Paler, L., 193
Paluck, E., 263, 311
parallel trends assumption, 217
partners, 326–328,
Pennycook, G., 288
Perlis, R. H., 305
Phillips, J. H., 198
piloting, 330–333
pivoting, 338–340

placebo-controlled experiments, 279–283
point estimator, 97
Polavieja, J. G., 231
Population Average Treatment Effect (PATE), 310
potential outcomes framework, 44–45
power, 124–125, 127, 129, 132
preanalysis plans, 334–337; populated, 340–344
Prillaman, S., 374
principles: Analyze As You Randomize (AAYR), 107,
110–115; design agnostically, 22, 25, 50–51, 87,
112; design early, 23, 26; design for purpose, 22,
25; design holistically, 22, 24; design often, 23, 26;
design to share, 23, 26–27; plug-in, 107–108, 110
process tracing, 203–208; design examples, 207–208;
illustration, 205

Q

qualitative research, 73, 85, 99, 103, 107, 203–208

R

Radl, J., 231
Rainer, I., 310
Rand, D. G., 288
randomization, 249; inference, 63–64, 113–15; in
sampling, 68–72. *See also* assignments; sampling
randomized experiments, 249–298
randomized saturation experiments, 288–291; design
examples, 291–292
reality tracking, 48–49
reanalysis, 351–355
reconciliation, 341–344
redesign, 16–17, 137–147; over answer strategies,
143; over estimators, 145; over two data strategy
parameters, 141; under model uncertainty, 139
regression discontinuity design, 221–226; design
examples, 226; illustration, 223
replication, 356–358; archive, 349–351
Revkin, M., 207, 241
robustness, 113, 123, 356
Root-mean-squared Error (RMSE), 123, 131
Rubinstein, A., 306
Rudkin, A., 312

S

Samii, C., 122, 216, 226, 293–294, 309, 360
Sample Average Treatment Effect (SATE), 11, 251
sampling, 65–67, 67–68; bias, 123; cluster, 70; clus-
ter randomized, 190–193; cluster randomized
design examples, 193; complete randomized, 69;
convenience, 72–73; multi-stage randomized,
70; nonrandomized, 72–73; purposive, 72–73;
for qualitative research, 73–76; simple random,
69–70, 185–190; simple random design exam-
ples, 189–190; simple random illustration, 187;
stratified, 69–70
Santillana, M., 305
Scacco, A., 278
Schwarz, S., 312

- Sejdinovic, D., 189
selection-on-observables, 208–211; design examples, 211
Siddiqui, N., 221
Simonson, M., 305
simulation, 116–117, 125–126; error, 130–131; in MIDA, 117. *See also* Monte Carlo simulation
Slough, T., 323
Snyder, J., 208
Stable Unit Treatment Value Assumption (SUTVA), 251
statistical power. *See* power
Stecula, D., 190
stepped-wedge experiments, 283–288; design examples, 288
Stokes, L., 221
Stokes, S., 193
structural estimation, 305–310; design examples, 309–310
subgroup designs, 263–266; design examples, 266
success, 15,17,19,123,331
Swire, B., 266
- T**
Tausanovitch, C., 198
Teele, D., 323
treatment assignment. *See* assignment
Trebbi, F., 310
Tillery, A., 337
true standard error, 123, 129
trust game, 242–248; design examples, 248
two-arm randomized experiment, 250–257; DAG Illustration, 252; examples, 257
two-way fixed effects estimator, 214–215
Type I error, 102, 123
Type II error, 102, 120, 123
Type-S error rate, 116, 123
- U**
uncertainty, 94–95; Bayesian, 94; frequentist, 95
UTOS framework, 53
- V**
value for money, 124, 329
variable, latent, 198–202
- W**
Walker, M. W., 292
Warren, S. S., 278
Warshaw, C., 198
Weintraub, M., 278
Westwood, S., 248
Wilke, A. M., 271, 283, 306
Wood, E., 323
writing, 344–346
- Y**
Yemane, R., 231
Young, L., 323
- Z**
Zelizer, A., 298
Zingher, J. N., 298