Declaring and Diagnosing Research Designs

A Book Proposal to Princeton University Press

Graeme Blair[†]   Jasper Cooper[‡]   Alexander Coppock[§]   Macartan Humphreys[††]

December 19, 2017

1   Proposal

Declaring and Diagnosing Research Designs will:

1. Introduce a framework for conceptualizing designs in a demonstrably *complete* way

2. Guide readers in the formal declaration and interrogation of research designs

3. Provide a library of standard and complex research designs commonly employed in the social sciences, which readers can adopt or modify for their own research

4. Use declared research designs to highlight design tradeoffs and illustrate major principles of research design

The book builds on two ideas. First, designs can be "declared" by formally stating design features in the proposed *<Model*, *Inquiry*, *Data Strategy*, *Answer Strategy>* framework (MIDA). Second, once declared, designs can be "diagnosed." Design diagnosis makes it possible to investigate the quality of the answers provided by a design under a variety of conditions.

The core idea of formally declaring a design is that a design can be treated as self-contained object that can be interrogated. Once thought of as an object in this way a design can be shared, modified, and used. Our goal is to alter the way readers think

[†]Assistant Professor of Political Science, UCLA. graeme.blair@ucla.edu. `https://graemeblair.com`

[‡]Ph.D. candidate in Political Science, Columbia University. jjc2247@columbia.edu. `http://jasper-cooper.com`

[§]Assistant Professor of Political Science, Yale University. alex.coppock@yale.edu. `https://alexandercoppock.com`

[††]Professor of Political Science, Columbia University. mh2245@columbia.edu. `http://www.macartan.nyc`

about research designs, whether or not they end up using our accompanying tools. We hope that the framework will support research transparency, facilitate design development and evaluation, and help structure critique.

2   Format

We propose that the book be published as a hardcopy text-treatise and accompanied by a comprehensive set of open-access online materials. We hope the online materials will make the book more valuable and vice-versa. The online materials would include:

1. A complete online version of the book, organized as a website, a development version of which can be found at `http://DDRD.DeclareDesign.org`.

2. Interactive versions of most visualizations included in the physical book. An example would be a power curve that shifts depending on design parameters.

3. A design library that will be continuously updated and expanded. We will encourage readers of the book to upload their own designs.

In addition to the online materials, the book will be accompanied by a set of three software tools:

1. The `DeclareDesign` package, an `R` package which is used throughout the book to declare and diagnose designs

2. The "Design Declarer" web interface that allows readers to generate and diagnose designs without using `R`

3. The "Design Inspector" web interface that allows readers to read, modify, and use designs without using `R`, including those in the design library

The online materials will integrate these applications naturally, coupling text and figures with interactive computational tools that let readers explore the properties of research designs in real time.

3 Motivation

Advances in causal analysis from the graphical models and potential outcomes frameworks have revolutionized empirical social science research. In recent years, focus has shifted to more careful definition of causal quantities of interest and a more direct focus on study design as a basis for inference.

Experimental research in particular has experienced enormous growth in the past two decades. For many years, whole disciplines were viewed as non-experimental sciences; this is no longer the case. Likewise, observational quantitative research has shifted from wholesale modeling of outcomes on the basis of large sets of predictors to explicitly seeking to approximate randomized experiments through the use of matching, regression discontinuity, and instrumental variable designs. The insights from the "identification revolution" have also illuminated descriptive research practices — not least by clearly delineating how causal questions differ from descriptive questions. More striking perhaps is the growth in the use of experimental techniques as measurement devices. Rising attention to causal identification is also visible in some qualitative research practices, with a move towards defining inferential goals as causal estimands, articulating mappings from data to inferences, and in some cases pre-registering qualitative strategies.

Researchers, whether focused on observational or experimental approaches, are increasingly paying close attention to the properties of research designs and the ability of these designs to justify inferences. In doing so, they face two challenges.

First, surprisingly little attention has been paid to the fundamental question of what constitutes a design. This lack of clarity carries risks both before and after the implementation of a study. If designs are incompletely specified ex ante, it is difficult for researchers to assess their strengths and improve them. If designs are incompletely specified at the time of analysis, researchers may choose inappropriate analysis procedures or worse, may only report the most attractive model from a set of possible specifications. If designs remain unspecified after analysis, it may be difficult for other scholars to replicate a study or to judge whether a given type of reanalysis is justified.

A second problem compounding the first is the dearth of tools for assessing design properties comprehensively. At one extreme, a minority of advanced scholars conduct fully-fledged simulations requiring programming skills beyond those of many applied researchers. At the other, researchers resort to basic power calculators that cannot capture important features of the research setting, such as the sampling or assignment strategy. We know of no readily-available methods to investigate whether a design will be biased, have good coverage, or meet other important inferential criteria beyond power.

DDRD addresses these two problems. It introduces a framework to describe the distinct elements of a research design. It provides a library of common designs, fully characterized. These designs elucidate 50 general principles of research design.

4 The Book

The book is organized in four main sections as described in Box 1 below. We describe the goals of each section in turn, and present a detailed (draft) plan in Appendix A.

Box 1

Contents

A : Declaring and Diagnosing Research Designs

  1 MIDA, in principle, in practice [20 pages]
  2 Diagnosing research designs [8]
  3 How to read this book [5]

B : Design Library

  4 Observational designs for descriptive inference [30]
  5 Experimental designs for descriptive inference [30]
  6 Experimental designs for causal inference [30]
  7 Observational designs for causal inference [30]

C : Principles of Research Designs

  8 Model principles [30]
  9 Inquiry principles [50]
  10 Data strategy principles [60]
  11 Analysis strategy principles [60]
  12 Cross-MIDA principles [40]

D : Putting Declared Designs to Use

  13 Researchers: Design development and registration [8]
  14 Peers: Better scholarly critique [8]
  15 Funders: Evaluating and supporting research [6]

[415 pages total, approx.]

4.1 The MIDA Framework

Part A introduces the MIDA framework for characterizing research designs, which builds on ideas about research designs in King, Keohane and Verba (1994). The framework conceptualizes a design as comprising four elements $< M, I, D, A >$:

1. A **model**, M, of how the world works. The model specifies the moving parts — the variables — and how these are causally related to each other. In this sense the model provides the context of a study, but also a speculation about the world.

2. An **inquiry**, I, about the distribution of variables, perhaps given interventions on some variables. In many applications I might be thought of as the "estimand." Some inquiries are statements about the values of variables, others about the causal relations between variables. In all cases however the inquiry should be answerable given the model.

3. A **data** strategy, D, generates data on variables. Note that implicitly the data strategy includes case selection, or sampling decisions, but it also represents interventions such as assignment of treatments or measurement strategies. A model M tells you what sort of data you might observe if you employ data strategy D.

4. An **answer** strategy, A, that uses data to generate an answer.

MIDA is deliberately pitched at a high level of generality so that it can be relevant for research whether it seeks to answer causal or descriptive questions, uses experimental or observational data, or employs quantitative or qualitative inference strategies.

A key feature of this bare specification is that if M, D, and A are sufficiently well described, the answer the design provides to question I has a well defined distribution. Moreover, one can construct a distribution of comparisons of the answer provided by the answer strategy (a^A) to the correct answer under M (a^M).

The ability to calculate distributions of answers, given a model, opens multiple avenues for assessment and critique of research strategies. How good is the answer you expect to get from this strategy? Would you do better with a different data strategy? With a different analysis strategy? How good is the strategy if the model is wrong in some way or another?

We call this *design diagnosis*. The key idea of design diagnosis is that if one has a fully specified design one can predict what will happen when a design is implemented, at least conditional on the assumptions made explicit through declaration. A design is "fully-specified" if it can be simulated. From there it is an easy step to assess what would result from multiple implementations of a design and so build up a distribution of quantities of interest — what we term "diagnosands." Researchers are familiar with the idea of

5

statistical power — the probability that a design will produce a *p*-value below some specified cutoff. That same idea can be extended to many properties of a design: once specified, repeated iteration of the design reveals its bias, the precision of the estimates, the performance of one design relative to another in a given setting, and so on.

Alongside the introduction of the core ideas of design declaration and diagnosis, Part A presents the `DeclareDesign` package and web applications, describing how readers can characterize designs using inbuilt functionality or external tools.

4.2 Design Diagnosis

Part B introduces a design library. The characterization of designs we advocate allows for arbitrarily complex and idiosyncratic designs. Many studies however use quite standard setups. In this section we provide a characterization of 25 common designs, describing their underlying assumptions and basic properties. Defined as objects early in the book, these same designs can be reused or modified to explore design principles later on.

Each design is: (a) introduced in terms of the MIDA framework; (b) formally characterized in the text; (c) illustrated; and (d) accompanied by short set of exercises inviting the reader to consider non-obvious features of the designs. In addition, the text points to examples of where these designs have been used as well as to possible applications to other areas of research.

4.3 Principles

Part C introduces general principles of research design. We use the framework from Part A and the designs declared in Part B to introduce a set of approximately 50 principles of research that cover data gathering, question formation, and data analysis. We seek principles that have general import — that is, that may be relevant for any research strategy — but that are under-appreciated.

We believe the strength of the book is in the ability to use the framework developed in earlier sections to show how these principles matter for specific research designs and how modifiable details of the designs can make issues raised more or less salient.

In the accompanying sample we provide examples of these principles and in the detailed table of contents we provide a draft listing of 50 possible principles. These are all principles that have (a) general import even if they are not broadly adhered to and (b) are well illustrated by design declaration — that is, declaration and diagnosis of designs shed light on when concerns raised by these principles are more or less important.

Each of these principles will be accompanied by (a) a design; (b) a figure illustrating the principle; and (c) short exercises that challenge readers to explore the principles further or in other domains.

4.4 Putting Declared Designs to Use

The final part of the book is the shortest part and points readers to broader uses of the book and framework, beyond the central role for researchers seeking to develop strong designs. The section has three chapters focused on three different audiences.

The first chapter, aimed at researchers, discusses the linkages between design declaration and pre-registration. The advance characterization of a design can greatly facilitate the registration process and in this section we describe how this is best done.

The second chapter focuses on the role that design declaration can play in scholarly critique. A common approach to scholarly critique is to alter some feature of an analysis strategy and to assess how results change. The problem with this approach is that when divergent results are found, third parties do not have clear grounds to decide which results to believe. A more coherent strategy facilitated by design declaration is to learn about what a study *could have* revealed under different analysis strategies, not just what the original author, or a critic, reports *was* revealed. Answering this question provides grounds for supporting alternative analyses on the basis of the original authors' intentions and not on the basis of the degree of divergence of results. Conversely, it provides authors with grounds to question claims made by their critics.

The third chapter focuses on research funders. We believe that design declaration can alter the way that research designs are evaluated. Currently funders of research have limited tools for assessing research quality, relying largely on the reputations of researchers. Sometimes power analyses are requested though these are often difficult to appraise. In contrast, access to declared designs allows funders to directly interrogate a design, to check whether the design will provide reliable learning under different conditions or if operated at different scales, under assumptions provided by researchers. This chapter describes how such review and evaluation activities can be conducted with declared designs.

5 Relation to Existing Work

Many excellent textbooks deal with particular features of research design. *DDRD* differs from these broadly in terms of (a) the integrated approach taken to conceptualizing design, rather than (only) a collection of tools; and (b) its hands-on approach to developing and diagnosing designs.

Integrated framework. Many design books focus on particular dimensions of research designs; many on analysis strategies, some on data collection strategies, such as sampling or assignment strategies. *DDRD* differs from these in using an integrated approach where data strategies are assessed in light of analysis strategies and vice versa. In this sense it is closer to books on design such as Gerring (2011) and Brady and Col-

lier (2010), although it takes a more formal approach. We note that some books take an even more expansive view of research design. Glennerster and Takavarasha (2013) for example includes sections on logistical details of field research such as engaging with partners, for example. Our book focuses more narrowly on the analytic properties of designs, though doing so in a more formal and technically-integrated way.

A difference with many design books is that *DDRD* is not comprehensive of a particular set of strategies, and it is not intended to be. Although many important workhorse designs used in social sciences are covered, the goal is not to cover all possible designs. The set of principles illustrated are somewhat eclectic. The reason for this is that the book seeks less to deliver a set of results and more to provide researchers with a strategy and a tool to address unanticipated design questions.

Hands on approach. *DDRD* is also hands-on in the sense of developing principles in an applied context. For this, it shares with many other books a strategy of integrating learning about general principles with the use of statistical software. These include Gelman and Hill (2006) and James, Witten and Hastie (2014), though both of these focus primarily on analysis methods. *DDRD* instead focuses on designs in an integrated way — not as a checklist, but as a collection of relations between objects that need to be mutually coherent. In this sense, it is close to books such as Gerber and Green (2012) and Dunning (2012) that have a goal of helping researchers use causal models to think about the inferences that can be made from different sorts of data generating strategies and insisting on the importance of maintaining consistency between data generation strategies and analysis strategies. It differs from these in its more practical focus, aimed at helping researchers develop complete designs and validating them, as well as linking general principles to specifics of designs.

Like the classic text King, Keohane and Verba (1994), *DDRD* takes a holistic view of research design: our framework can be applied to empirical social scientific work in many traditions, including most qualitative and quantitative approaches. We build on King, Keohane and Verba (1994) by applying insights from the causal inference "revolution" across a wide variety of research settings.

In all we believe that *DDRD* occupies an unique place among current books on design, offering a generally applicable framework for thinking about designs in an integrated way and a set of practical tools to develop designs and learn about their properties. We know of no other book that seeks to do this.

6 Audience

DDRD is not intended as a standalone textbook for a course. Instead, it serves as an accompanying text, which could be used in a wide variety of courses, and as a reference for researchers and funders. Likely courses at institutions that *DDRD* could accompany include:

1. General research design courses in masters and Ph.D. programs in political science, public policy, sociology, and business schools

2. Experiments and causal inference courses in Ph.D. programs in political science, public policy, and sociology

3. Advanced experiments and quantitative methods courses at top undergraduate institutions

4. Applied substantive courses in the social sciences that review complex empirical work

For these classes the book could serve as a set of modules, focusing students on fundamentals of research design, or it could be used to accompany other statistical classes, providing tools for students to explore the properties of any model that is introduced. Beyond methods courses, many graduate courses in the social sciences focus on published research with highly complex protocols and methods. The findings of such research are difficult to interpret without some understanding of the underlying design. Our short-format overviews could be given as handouts to provide an introduction to the underlying logic and tradeoffs of the methods employed.

Beyond the classroom the book is aimed at evaluation practitioners — impact evaluation and measurement and evaluation (M&E) specialists in government, international lending institutions such as the World Bank, bilateral aid agencies, major donors, and political consulting organizations such as the Analyst Institute — who all support research but have few tools to assess the integrity of research proposed to them.

7 Timeline

We expect to complete the book over the next 18 months, with research and programming assistance funded in part by a grant from the Laura and John Arnold Foundation.

We have some milestones set for the manuscript development:

- Graeme Blair will teach a course around this content at UCLA in Winter 2018 and 2019 using 15 declared designs and 15 elaborated principles

- Alex Coppock will teach a course around this content at Yale in Spring 2018 and 2019 using 15 declared designs and 15 elaborated principles

- By October 2018, we will complete a draft of the book material

- We will then complete the remainder of the manuscript by April 2019

8 Enclosures

We enclose:

1. Sample pages containing two subsections of Part B ("Simple Random Sampling," "Regression Discontinuity Designs") and one subsection of Part C ("Questions Should Have Well-Defined Answers"). The principle we provide as an example for Part C is one that emerges naturally from the integrated approach to design development we describe; in standard approaches when estimators are defined separately from estimates it can be easy to implement analysis without noticing that the estimand itself is not well defined.

2. A prototype of the e-book, including interactive applications that interrogate various research designs. This prototype can be viewed at `http://DDRD.DeclareDesign.org`; an example of the inspector application integrated into the book can be seen at `http://DDRD.DeclareDesign.org/simple-random-sampling.html`.

3. A draft detailed table of contents (Appendix to this document)

4. An unpublished article manuscript outlining the framework to be expanded upon in the book, available at `http://declaredesign.org/paper.pdf`

A   Appendix: Detailed (Draft) Table of Contents

1. Part A: Declaring and Diagnosing Research Designs

   1.1. MIDA, in principle, in practice

   1.2. Diagnosing research designs

   1.3. How to read this book

2. Part B: Design Library

   2.1. Observational designs for descriptive inference

      2.1.1. Simple random sampling

      2.1.2. Stratified clustered random sampling

      2.1.3. Measuring latent variables

      2.1.4. Topic analysis

      2.1.5. Respondent driven sampling

      2.1.6. Multilevel regression and poststratification

   2.2. Experimental designs for descriptive inference

      2.2.1. Audit experiments

      2.2.2. List experiments

      2.2.3. Randomized response

      2.2.4. Conjoint experiments

      2.2.5. Experimental games

   2.3. Experimental designs for causal inference

      2.3.1. Multiarm designs

      2.3.2. Experiments with blocks and clusters

      2.3.3. Factorial designs

      2.3.4. Encouragement designs

      2.3.5. Stepped wedge designs

      2.3.6. Crossover designs

      2.3.7. Parallel design for mediation effects

      2.3.8. Partial population design for spillover analysis

      2.3.9. Selective trials

      2.3.10. Adaptive trials

References

Brady, Henry E. and David Collier. 2010. *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Lanham, Maryland: Rowman & Littlefield Publishers.

Dunning, Thad. 2012. *Natural Experiments in the Social Sciences: a Design-based Approach*. Cambridge: Cambridge University Press.

Gelman, Andrew and Jennifer Hill. 2006. *Data Analysis Using Regression and Multi-level/Hierarchical Models*. Cambridge: Cambridge University Press.

Gerber, Alan S. and Donald P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. New York: W.W. Norton.

Gerring, John. 2011. *Social Science Methodology: A Unified Framework*. Cambridge: Cambridge University Press.

Glennerster, Rachel and Kudzai Takavarasha. 2013. *Running Randomized Evaluations: A Practical Guide*. Princeton: Princeton University Press.

James, Gareth, Daniela Witten and Trevor Hastie. 2014. *An Introduction to Statistical Learning: With Applications in R.* New York: Springer.

King, Gary, Robert O. Keohane and Sidney Verba. 1994. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.