1. Introduction to the *Handbook of Research Methods and Applications in Comparative Policy Analysis*

B. Guy Peters and Guillaume Fontaine

1. INTRODUCTION: SUBSTANCE AND METHODS IN COMPARATIVE POLICY ANALYSIS

Comparative policy analysis (hereafter CPA) is an emerging area in social sciences and already a major one in policy studies. As of December 31, 2018, the Web of Science's database registered 243 references of peer-reviewed articles, whose authors used "comparative policy" as descriptor (as key words, in the title or in the abstract). This is still less than "comparative politics" (776 references during the same period), but it is much more significant than "comparative public administration" (39 references), considering the same token. More important, though, is the upward trend of the literature explicitly referring to CPA as a descriptor of the substance and method of a research, as more than 56% of these titles were published after mid-2015.[1]

Given the multi-disciplinary nature of policy studies, the methodological challenges in CPA are arguably more complex than for other research areas in social sciences. Therefore, there is a greater need to provide an overview of the available methods, as well as their strengths and weaknesses for addressing theoretical issues like policy change, time and context in policy design, and so on. Hence, this handbook of methods for comparative policy analysis is intended to advance the understanding of methodology in the study of comparative public policies, as much as to broaden the array of methods and techniques considered by CPA scholars in their research design. The methods discussed in its 22 chapters, including this introductory study, consider not only the nature of public policies but also how to deal with the complexities of policy across political systems and across policy domains and across time. These methods are also of interest for practitioners, who face many policy problems of international dimensions and have a lot to gain from policy learning and policy diffusion, but often lack the methodological tools to deal systematically with the comparison of most-similar and most-different systems.

The need for such a handbook has become especially pressing for two additional reasons. The first is the increasing diversity and complexity of available methods in social sciences, which makes their selection a challenging exercise for policy design in research and practice, beyond substantive issues. Some are of relatively recent vintage and have been adding significant analytic power to CPA, like the qualitative comparative analysis methods (QCA) (see **Chapter 15** by Thomann). Others have been common in some disciplines but are now becoming more and more popular among policy scholars and practitioners, like process tracing (see **Chapter 16** by Fontaine), focus groups (see **Chapter 17** by Marier, Dickson, and Dubé), and ethnography (see **Chapter 18** by Pacheco-Vega). Still others are actually decades old but have

1

recently experienced a new appraisal as outstanding quantitative and qualitative techniques in the social sciences, like text-as-data analysis (see **Chapter 12** by Gilardi and Wüest) and the Q-methodology (see **Chapter 19** by Molenveld).

The other compelling reason for considering the range of methods available to CPA scholars is the necessity to overcome the limits of quantitative techniques (especially standard linear regressions and econometric models), which are sometimes mistakenly identified as the best way (if not the only effective one) to conduct cross-case, cross-area, or longitudinal comparative analysis. The overwhelming publication of probability-driven research outputs in high-impact journals – sometimes assimilated to grounded theory or to knowledge-based policy making – may produce the wrong impression that contemporary social sciences, in general, and CPA, in particular, rely on a limited range of methods considered to be valid. Indeed, probabilistic methods are generally used with little consideration of available alternatives, because of the assumptions that are involved in the regression or model designed by scholars for theory-building or testing. Moreover, they are often used in a deterministic fashion, following an oft-cited motto in public management "if you can't measure it, you can't improve it".

We consider this reliance on a limited range of methods unfortunate for three major reasons. First, it affects the research agenda in CPA by neglecting major theoretical and practical issues for policy analysis and design: focusing on a limited number of methods will also limit the range of issues to be taken into consideration. Yet not all research questions are amenable to the same methods (see our typology of methodologies below). Second, it is often desirable to have results using one method compared with those from an alternative method, in order to gain additional insight into a research question (Webb et al. 1999). Taken as a whole, this handbook demonstrates that mixed- and multi-methods can provide a more complete and nuanced account of causal relationships (see in particular **Chapter 13** by Dunn and Peters). Third, understanding policies and policy choices requires detailed examinations of those individual policies and their consequences, which often implies conducting case studies or small-N comparison (see **Chapter 14** by Beach).

A more open-minded and conscious selection of methods and techniques to address a research or a policy problem should improve the quality of CPA. Hence spreading our methodological net as wide as possible is likely to provide more robust insights. Based on these premises, this introductory study raises two central questions that run throughout the handbook and are attended differently by the authors: What are the best available methods to conduct systematic CPA across time, space, and areas? And how can scholars and practitioners select them or combine them to improve the internal consistency and the external coherence of policy design in research and practice?

The remainder of this chapter is divided in four sections. Section 2 presents four major theoretical issues currently under discussion in CPA, in order to show the methodological implications of theory-driven research and how these can be attended by the methods presented in this handbook. It considers the structure versus agency debate, the study of policy change and the role of context and time. Section 3 presents the methodological issues based on a sample of 80,000 peer-reviewed articles from the Web of Science. It describes the methodological shift in social sciences and the methods preferences by areas thereof. Section 4 draws on the research design and methods selection stemming from the former sections. It builds on a typological model to deal with the diversity and complexity of these methods, and then it proceeds with comparing the scope and limitations of extensive and intensive research designs

in outcome explanation or interpretation. Eventually this section underlines the difficulties raised by case selection in a within-case study or a small-N comparison. Section 5 draws on some provisional conclusions that stress the importance of methods for comparative policy research and practice and presents the handbook's organization.

2. THEORETICAL ISSUES

Many theoretical perspectives are relevant for CPA in one way or another (Peters and Zittoun 2016; Weible and Sabatier 2017). In a research design these theories need to be connected with methods that can enable scholars to confirm or to disconfirm their hypotheses. If CPA is to provide a cumulative knowledge to improve policy design in theory and practice, such methods need to be capable of testing them comparatively. In this section we address some of the central issues at stake in the contemporary debate, as related to their methodological implications.

2.1 Structure versus Agency

The question of how structure and agency are related in explaining outcomes is pervasive in social sciences in general, and in CPA in particular. How much does the institutional structure within which decisions are made matter, as compared to the influence of individuals who actually make these decisions? In contemporary political science and sociology this question is at the core of the debates between *methodological individualism*, favoring behavioral and rational choice theories, and *neo-institutionalism*, arguing for the causal determination of structure. The correct answer to the question, of course, is that both matter but that does not solve the problem of where scholars should begin their analysis.

The structure-versus-agency debate appears particularly important in policy studies, since a policy is essentially made within institutions (see **Chapter 22** by Kay). That being said, there is undoubtedly a danger in anthropomorphizing institutions, thereby ignoring the nature of the individuals who occupy positions within those institutions (Scharpf 1997). The bottom line is that individuals make decisions that result in policies being adopted and implemented, but they do not act in a vacuum. They are influenced by the institutions within which they work, and with which they interact. Therefore, we would assume (with good cause) that different institutions will process policies differently and will privilege certain types of policy choices.

Thus, thinking about only one of these approaches to explanation is likely to provide only a partial vision of what is occurring when the policy is made. As already noted, this theoretical problem is endemic in the social sciences but the importance of institutions in a policy area – legislatures, bureaucracies, interest groups, etc. – makes the problem more acute. The extent to which individual decisions and the norms and information held by institutions may be intertwined is difficult to untangle, and may well require qualitative methods to identify the premises on which decisions are made.

2.2 The Role of Context

The status of context in comparative politics is somehow paradoxical. On one hand, the methods used for comparative analysis sometimes attempt to eliminate those contextual

factors when focusing on the relationship between an independent and a dependent variable. In particular, the comparative method as described by Arend Lijphart attempts to hold constant as many contextual factors as possible in order to assess those relationships (see **Chapter 2** by Peters). On the other hand many theories, especially those derived from comparative politics, attempt to involve contextual factors directly in the analysis of policy, through most-similar or most-different systems research designs (see **Chapter 3** by Anckar).

This seemingly paradox points to the importance of within-case and small-N comparison in theory-testing (see **Chapter 4** by Dowding). Within-case analysis allows for a full exploration of the contextual factors that influence a policy area, and that detailed analysis is not "mere description" (Gerring 2010), but does enable a full understanding of the dynamics of policy. But the comparative method as framed by Lijphart is concerned with cross-case analysis, and developing generalizations from multiple cases. In that sort of analysis, the context – as important as it is – may produce spurious results and therefore needs to be controlled in the research design.

One of the virtues of CPA is that it brings context directly into the analysis. Policy making does not occur in a social, cultural, or economic vacuum: it is influenced – and perhaps decisively so – by environmental factors. This is true even when experimental designs for research are being utilized: the context of the experiment may still matter, and should be included in the conceptualization of the research project (see **Chapter 9** by John). Some methods discussed in this handbook emphasize culture and context when attempting to assess the narratives elaborated by a policy area's actors (see **Chapter 20** by Smith-Walter and Jones) but those factors are also important for any study of policy. For example, there is a significant body of literature arguing that economic factors are much more important than political ones for determining public spending. It is crucial to understand such factors as political culture and social structure when attempting to understand policy choices, and the effects of those choices within a society.

2.3   The Study of Policy Change

To a great extent the study of contemporary public policy is the study of policy change, given that there are few virgin areas into which governments can intervene (Carter 2012; Hogwood and Peters 1987). Most policy domains are occupied by a host of policies, so policy making basically consists in changing those existing policies and only occasionally fitting some new policy into that crowded domain. Consequently one of the main methodological challenges for CPA is coping with change.

A major modality of policy change is diffusion and learning (see **Chapter 8** by Porto de Oliveira). One of the virtues of thinking comparatively about public policy is that it rapidly becomes apparent that the world is a laboratory for policy change, where individual policies are tantamount to natural experiments. Policy innovators and activists are at work in many countries attempting to develop better solutions for common problems, and those ideas then are diffused across countries. There is an extensive literature on learning and diffusion (Dunlop and Radaelli 2013; Rose 1993) that points to the importance of understanding how policies travel across political systems, as well as the difficulties in making those transfers in practice. This interest in diffusion and learning has increased with the concern for "evidence-based policy making" and the ability to utilize evidence from one setting to inform policy design in

others (Cairney 2016; Pawson 2006), despite the inherent difficulties in translating the success in one setting to another.

Among all the dimensions of policy change, the most obvious is that governments who decide to change existing policies have to proceed differently according to the political systems in which they are embedded. For example, there is an extensive literature on the welfare state retrenchment among the developed democracies that provides important examples of how to understand that kind of policy change (Castles 2004; Starke 2006; Van Kersbergen and Vis 2014). That change may be the result of political change, policy innovation, policy entrepreneurship, or even the whims of a powerful individual (agency). To study this process requires measuring the nature of the policy at one point in time (either by quantitative or qualitative means), measuring the extent to which that measurement has changed, and then attempting to understand the difference (see **Chapter 10** by Tosun and Schnepf).

Once again, CPA should be especially sensitive to the contextual issues involved in utilizing evidence from policy making across cultures and societies. Some indicators of a public policy (e.g. public expenditures) may be relatively simple and usable across cultural boundaries, but many others are embedded in their social and cultural environments (see **Chapter 11** by Erkkilä). This list of concerns about CPA could be extended, but the bottom line is that identifying and measuring the variables involved is not as easy as for voting behavior, or public opinion.

2.4   The Problem of Time

Stemming from the problem of policy change, time is of the essence in CPA. The effects of a policy may be spread across time, hence requiring some consideration of the long- and short-term consequences of action (see **Chapter 7** by Jaramillo). Although most explicitly assumed CPA regards cross-case analysis, within-case longitudinal comparison is an equally important form of analysis. One could, in fact, argue that it is more important because it holds constant many possible confounding factors – the context mentioned above – and therefore allows for a clearer understanding of the relationships among variables in the analysis.

Yet this raises a twofold problem: a methodological one and a theoretical one. On one hand, techniques such as time-series and pooled time-series attempt to identify processes of development and change within a policy and relate them to changes in the presumed independent variables. But as powerful as studying policy across time can be, it also raises significant methodological challenges (see Kellstedt and Whitten 2013). The most common statistical problems met in CPA to assess the necessity and sufficiency of a condition are related to omitted variables (which are virtually infinite in open systems) and multi-causality (i.e. competing theories of causation of a single phenomenon). Other issues require building complex econometric models, to deal with spuriousness (i.e. a variable affecting both the cause and the effect), equifinality (i.e. different independent variables causing a similar effect on a dependent variable), circularity (i.e. independent variables affecting one another), or heteroscedasticity (i.e. a variation among variables from the same set).

On the other hand, time does imply change in the context, and can be a significant source of invalidity in findings (Shedish, Cook, and Campbell 2002). Time, in a simple chronological sense, may be important because of external changes, but also because of changes within the subjects of the research (maturation) and changes in the researchers themselves (instrumentation). Dealing with these sources of invalidity is paramount both for qualitative

and quantitative research. Moreover, time-series analysis may explain very high correlations among factors in regressions, simply because the observations are not independent. There are statistical techniques to address this problem inherent to longitudinal comparison, but those statistical methods used to eliminate some of the effects of time may also undermine the understanding of some of the important effects of change across time. For qualitative researchers the effects of time will be more difficult to overcome.

Eventually, the difficulties associated with time in CPA may lead some scholars to take a more static approach to the phenomena they are interested in. But an assumption that variables, at any one point in time, do not have a past is likely to be fallacious. This produces another research question of where to draw the line when considering the influence of history. To understand contemporary policy in France, do we need to go back to World War II, to the Revolution, or to the *ancien régime*? The answer may depend on the particular policy question under study, and then again CPA scholars do need to understand the context. For what's left, understanding the background will illuminate the decisions being made at the present time, for any policy question.

In a nutshell, contemporary debates in CPA about structure and agency, policy change and the role of context and time call for different research designs, combining qualitative and quantitative methods. But how are scholars to combine these techniques in their research design? The following section provides some substantial elements based on bibliometrics to answer this question.

3.   METHODOLOGICAL ISSUES

CPA shows some particular features that justify bigger methodological concerns than in other research domains from the social sciences. The most important one is that the field is at the confluence of various disciplines, which have been developing their own strategies of description, explanation, and interpretation. In this section we characterize what we call "the methodological shift" in social sciences and we picture how this transformation has spread across many different research areas which are of interest for CPA.

3.1   The Methodological Shift in Social Sciences

Like any other field of research, CPA must make fundamental choices about the most appropriate means of developing evidence. But in addition to the standard concerns with methods, whether coming from within disciplines or from the confluence of disciplines, CPA itself may pose additional challenges to the prevailing methodologies, the most important of which is the complexity of the subject matter being studied. This is why some analytical frameworks have been more successful than others in explaining policy choices, either focusing on the role of advocacy coalitions (see **Chapter 5** by Nohrsted, Weible, Ingold, and Henry) or long-standing trends in the agenda-setting process (see **Chapter 6** by Chaqués Bonafont, Green-Pedersen, and Seeberg).

The discussion concerning the proliferation of methods – hereinafter referred to as the methodological debate in CPA – partly reflects the contemporary concern with methods and techniques in social sciences. As of December 31, 2018 the Web of Science's database registered almost 80,000 peer-reviewed articles claiming to use one or more methods described in

the present handbook.[2] This debate was sparked off during the heyday of comparative politics, political economy, and historical sociology, back in the 1970s. But the methodological shift actually goes back to the early 1990s, when the total number of articles assuming explicitly at least one of these research methods surpassed 100 per year. Another threshold was reached in 2002 with more than 1,000 articles published in a single year, and another one in 2017 with 10,000 articles (WOS 2019). It is noteworthy that more than half of these references were published between 2014 and 2018.

If we take political science, economics, and sociology as the "heartland" of CPA, then existing methodological discrepancies among scholars may not be too extreme. As a matter of fact, a majority of journals specialized in these three disciplines largely praise quantitative techniques (especially standard linear regressions) and experiments for theory-building and testing (Bardsley 2010). The ongoing debate, within these disciplines, between advocates of quantitative, qualitative, and multi-methods, reached a tipping point with the seminal book by King, Keohane, and Verba (1994). Yet, even if these authors and others (Brady and Collier 2010; Goertz and Mahoney 2012; Seawright 2016) state that the standards for quantitative and qualitative research are the same when it comes to theory-driven research, they undoubtedly show a bias towards a probabilistic approach to causation and causality.

This strategy has been challenged by scholars interested in institutional and cultural influences, arguably more in political science and sociology than in economics (Peters 2013). Some argue in favor of a configurational approach to causation through QCA, which has quite recently become popular in CPA (Engeli and Rothmayr Allison 2014; Ragin 2008; Rihoux and Grimm 2006). Others value the contribution of case study and small-N comparison as complementary qualitative methods (Beach and Pedersen 2016; Blatter and Haverland 2012). Still others advocate for set-theoretic research designs based on multi- or mixed methods (Berg-Schlosser 2013; Blatter and Haverland 2014).

That being said, those are by no means the only three contributing disciplines to CPA. Many others bring in complementary approaches, hence providing new insights on causality and causation. For instance, anthropology brings in a range of observational techniques – like focus groups, ethnography, discourse analysis, etc. – that rely more on the locus of the individual researcher than on existing quantitative indicators. Further, both law and history work more with documentary evidence through congruence analysis, process tracing, and other qualitative techniques. Last but not least, substantive policy domains – such as public health, international relations, environmental sciences, or even engineering – have made original contributions to the debate, to be integrated with those of the conventional approaches in social sciences. These alternative sources of evidence contribute to the development of CPA, even if they can also be a source of confusion and conflicting elements for the coherence and consistency of a research design, as will be discussed below.

3.2   Methods Preferences by Research Areas

The diversity of methods in social sciences and CPA makes it worth a thorough review of their historical trends before discussing their contribution to research and practice. Based on our sample of peer-reviewed articles, the literature can be divided into four groups. The first one includes research papers based on case study, small-N comparison, or experiments, three methods that are widespread in social sciences, for they have been utilized since the 1970s. The second one consists in the research papers grounded in quantitative methods such as

surveys, standard linear regressions, and time-series, which have been increasingly used since the early 1990s. The third one is made of qualitative methods and techniques whose utilization has been increasing since the early 1990s, including ethnography, focus groups, comparative historical analysis, multi- or mixed methods, and the Q-methodology. A fourth group includes papers based on methods and techniques that already existed in the 1980s but became standard operating procedures in CPA during the 2000s, such as the comparative method, QCA, and process tracing. Scholars may combine various methods; therefore the same article can appear in different groups.

All these methods actually benefited from two major events affecting policy making. On one hand they were fueled by the exponential effects of innovations in information and communication technologies, from the design of the first statistical software packages available to the general public in the 1980s, and from the growing industry of big data in the 2010s. On the other hand they experienced a faster and wider diffusion across the academy and the practitioners community due to the intensification of globalization effects, including the emergence and the multiplication of transnational policy problems, the digitalization of knowledge, and the blossoming global community of policy scholars embodied in the International Political Science Association (created by UNESCO in 1949) and its offspring, the International Public Policy Association (born in 2016), not to mention all regional associations like the European Consortium for Political Research (ECPR, born in the 1970s).

Cross-area data comparison indicates scholars' preferences for certain methods and techniques according to their research area (see Appendix table).

Group 1 represents 64.81% of the overall literature. Two-thirds of the research based on case study and small-N comparison are concentrated in six areas: environmental sciences and studies (28%), public administration and management (10%), economics (8%), political science (6%), education (6%), and geography (6%). More than half of the literature using experiment is to be found in five areas: economics (15%), environmental sciences and studies (13%), computer science (10%), political science (8%), and public administration and management (8%).

Group 2 represents 11.66% of the sample. It is also concentrated in a few substantive areas, as two-thirds of the research based on quantitative techniques and methods focus on three areas – economics (38%), environmental sciences and studies (18%), and public administration and management (14%). Scholars using these methods are also relatively interested in political science (9%) and international relations (3%).

Group 3 concentrates 22.25% of the total. Methods used in this group are more evenly distributed across areas. First, two-thirds of the research using ethnography are concentrated in six areas – anthropology (18%), sociology (11%), education (11%), geography (9%), international relations (8%), and public administration and management (7%) – but this research design is also applied to environmental studies and sciences (4%), political science (4%), and linguistics (4%). Second, two-thirds of the research based on comparative historical analysis are to be found in five areas – political science (17%), environmental sciences and studies (17%), economics (14%), public administration and management (10%), and geography (8%) – but this method is also utilized in sociology (6%), international relations (6%), and history (5%). Third, most research using the Q-methodology is utilized in eight areas – education (16%), environmental sciences and studies (16%), linguistic (12%), communication (11%), political science (9%), sociology (9%), geography (6%), and international relations (6%).

That being said, some qualitative techniques from group 3 are still overlooked in a variety of research areas. Focus groups techniques basically focus on four areas: public administration and management (22%), health care sciences, health policy, and general medicine (21%), environmental sciences and studies (11%), and education (9%). Two-thirds of the research based on multi-methods and mixed methods are concentrated in five areas – health care and health policy (18%), public administration and management (17%), environmental sciences and studies (13%), education (11%), and international relations (5%).

Group 4 represents 1.3% of the total. This one is also concentrated in a few areas but, as already mentioned, this is a group of emerging techniques that barely sums up 1,028 papers, which makes it less representative. The comparative method is concentrated in five areas – public administration and management (16%), environmental sciences and studies (13%), health care and health policy (13%), political science (11%), and international relations (5%). Process tracing is concentrated in political science (48%), international relations (27%), and public administration and management (16%). Lastly, QCA is concentrated in four areas – political science (34%), public administration and management (23%), environmental sciences and studies (20%), and international relations (13%).

In a nutshell, the diversity of quantitative and qualitative methods across different research areas in social sciences confirms that there is no one best way to conduct a systematic CPA. The nature of the object of comparison and the challenge of coping with differences across political systems, policy areas, and historical contexts force researchers to consider different types of evidence, thereby giving up the single-method research design that might be sufficient in other fields. How can scholars then preserve the coherence of their research design when combining different methods? This question will be addressed in the following section.

4.   RESEARCH DESIGN AND METHODS SELECTION

While CPA may well follow the mainstream regarding methodological developments of the social sciences, scholars involved in this field may find some advantage in utilizing a broader array of methods to deal with the complexity inherent to public policies, and the need to understand the roles played by process and context in case selection. This requires solving the classical problem of alignment between ontology and methodology, which will guide scholars' election for an extensive or an intensive research design.

4.1   Aligning Ontology and Methodology

The definition of a strategy to answer a research question is a matter of methods and techniques, but the definition of a criterion to choose a method is a matter of methodology (Sartori 1970). A methodology is best understood as the product of a philosophical ontology (the relationship between the mind and the world) and a scientific ontology (the relationship between the status of empirical data and our knowledge about the world) (Jackson 2016).

This "stratified ontology" (Sayer 2000: 11–12), grounded in the distinction between the real, the actual, and the empirical, stems from the difference between the intransitive and the transitive dimensions of knowledge, that is to say between the physical processes and social phenomena in the world, on one hand, and the discourses and theories about the world, on the other hand. On one hand scholars' philosophical ontology can be dualist, if they conceive

Table 1.1   Four methodologies

		Scientific ontology	
		Phenomenalist	**Transfactual**
Philosophical ontology	**Dualist**	Neo-positivism	Critical realism
	Monist	Analyticism	Reflexivity

Source: Jackson (2016).

a gap between the mind and the world, or monist, if they conceive the world as a continuity of the mind. On the other hand, their scientific ontology can be phenomenalist, if their knowledge about the world is based exclusively on empirically observable data, or transfactual, if their knowledge about the world is also based on detectable but non-empirically observable data.

The stratified ontology supports a four-category typology of methodologies, that goes way beyond the somehow sterile opposition of quantitative versus qualitative methods (Table 1.1). Type 1 (neo-positivism) combines dualism with phenomenalism, type 2 (realism[3]) combines dualism with transfactualism, type 3 (analyticism) combines monism with phenomenalism, and type 4 (reflexivity) combines monism with transfactualism (Jackson 2016). On one hand, realists and neo-positivists share a dualist ontology, so they concur in that causal explanations of the world are possible. Conversely, analyticists and reflexivists share a monist ontology, so they concur in that only interpretations of the world can be provided. On the other hand, neo-positivists and analyticists share a phenomenalist ontology, so they concur that the scientific knowledge about the world is exclusively based on directly observable data. Conversely, realists and reflexivists share a transfactual ontology so they concur in that their knowledge is also based on detectable but non-directly observable data.

On one hand, types 1 and 2 share a dualist ontology, so they concur in that causal explanations of the world are possible. Conversely, type 3 and 4 share a monist ontology, so they concur in that only interpretations of the world can be provided. On the other hand, types 1 and 3 share a phenomenalist ontology, which means they concur that the scientific knowledge about the world is exclusively based on directly observable data. Conversely, types 2 and 4 share a transfactual ontology, which means they agree in developing scientific knowledge with detectable but non-directly observable data.

Aligning these methodologies with our methods in CPA requires a clear understanding of the logic of causation they support, as well as their implications for the aims of intervention, the modalities of evaluation, and the outputs for policy analysis and design (Table 1.2). According to this typology, type 1 research advocates for a logic of causation based on variable-oriented explanation, aimed at falsifying law-like theories and predictive models through the counterfactual analysis of observable data. Type 2 research favors case-oriented explanations to build and test middle-range theories and contingent patterns grounded in the transfactual analysis of non-observable data. Type 3 research would rather adopt a logic of causation based on a variable-oriented interpretation to build and test ideal-type theories and typologies through the counterfactual analysis of observable data. Type 4 research values case-oriented interpretation for critical theory-testing and narratives based on the transfactual analysis of non-observable data.

Classical examples of theories grounded in these methodologies are the punctuated equilibrium theory – a law-like theory – for type 1 (True, Jones, and Baumgartner 2007), the logic of appropriateness – a middle-range theory – for type 2 (March and Olsen 2006), the social construction of policy design – an ideal-type theory – for type 3 (Schneider and Ingram 1997), and

Table 1.2   *Methodological implications of different ontologies for CPA*

Methodology	Logic of causation	Aims of intervention	Modalities of evaluation	Outputs for policy design
Type 1: Neo-positivism	Variable-oriented explanation	Law-like theories falsification	Counterfactual analysis of observable data	Predictive models
Type 2: Realism	Case-oriented explanation	Middle-range theory-building and testing	Transfactual analysis of non-observable data	Contingent patterns
Type 3: Analyticism	Variable-oriented interpretation	Ideal-type theory-building and testing	Counterfactual analysis of observable data	Typologies
Type 4: Reflexivism	Case-oriented interpretation	Critical theory-testing	Transfactual analysis of non-observable data	Narratives

Source: Adapted from Jackson (2016) in: Fontaine, Medrano, and Narváez (2019).

the argumentative turn in deliberative policy analysis – a critical theory – for type 4 (Fischer 2003). Of all the reasons that have made these theories robust explanations or interpretations of policy processes, the most noteworthy here is that their authors proceeded with a careful alignment between ontology and methodology. This means that each one of these theories is bounded to a particular combination of philosophical and scientific ontology, which commands a particular conception of causation and causality, which in turn commands a particular combination of methods and techniques.

Therefore the four types of methodologies should be used as guidelines to secure the external coherence and the internal consistency of a research design, when it comes to the choice of a method or the combination of multi-methods. This leaves us with two different research designs – extensive or intensive – as explained below.

4.2   Extensive versus Intensive Research Design

Since they vary across the methodologies described above, the meanings of causality (a causal relationship between two entities or events) and causation (a hypothesis or a theory about a causal relationship) raise three intertwined problems: regularity, necessity, and symmetry. Regularity refers to a causation based on multiple observations, which is essentially a question of quantitative measurements. Hence it lies at the core of probabilistic methods and is generally associated with symmetry and correlation (the Holy Grail of statistics), which means that a positive or negative variation of a cause is correlated with the symmetric variation of its effect. Necessity refers to a causation based on a detailed account of the factors involved in a relationship, which is essentially a qualitative problem. It is at the heart of deterministic methods, which are generally concerned with the positive dimension of a causal relationship – the actual effect of a causal force – rather than with symmetry.

The major difference here is that establishing regularity in a causal relationship is about the mean causal effect based on as many observations as possible, while establishing necessity in a causal relationship is about the evidence that connects a cause to an effect based on a few cases or even a single-case study. Confirming or disconfirming regularity requires an extensive research design, while confirming or disconfirming necessity commands an intensive one (Sayer 2000). Both types of causation are actually complementary. For instance, based on a large-N

data panel we could observe a correlation between the event A (smoking) and the event B (dying of lung cancer), but this actually tells us nothing about how this particular cause produces this particular effect. Therefore a full-length theory of A causing B would require a multi-method research design combining an extensive approach (to establish the mean causal effect caused by A on B) with an intensive one (to establish the causal mechanism linking A to B).

Still, each part of such a research design aims at confirming or disconfirming different kinds of theories. An extensive research design may produce a causal law or a general theory. It aims at dealing with causality as a formal relation of similarity, which leads to asking which conditions are theoretically necessary to trigger a process. In open systems – such as public policies – this can only be done with semi-experiments (by manipulating the real world) or with statistics (in search of mean causal effects). Conversely, an intensive research may produce an instrumentalist law or a middle-range theory. It seeks to address causality as a substantial relation of connection, which leads to asking what it is about an entity that produces a causal power. This can only be done through deep within-case study or small-N comparison (in search of causal mechanisms).

The difference between intensive and extensive research designs does not overlap the dilemma between quantitative and qualitative methods (as a matter of fact they may both combine these methods). Yet intensive research is akin to case study analysis and small-N comparison, unlike extensive research which requires large-N comparison occasionally completed by case studies. In either case, multi-methods are utilized for exploratory purposes in theory-building and for confirming or disconfirming purposes in theory-testing. But only intensive research may be interested in deep within-case studies of non-representative processes.

Eventually, this division of labor refers to different approaches to the issue of theoretical generalization and different appreciations regarding the number of cases or observations deemed sufficient to build and test a causal explanation or interpretation. It is obviously risky for the coherence and the consistency of a research design to combine both intensive and extensive approaches in a single research design, without having previously determined what kind of causation they are expected to confirm or disconfirm. Therefore in practice, regularity is often grounded in the phenomenalism of types 1 and 3 methodologies, while necessity is more commonly grounded in the transfactualism of types 2 and 4.

Moreover, for types 1 and 3, the external validity of a causal relationship depends on the correlation between an independent and a dependent variable, so that the higher the number of observations, the better for the research. From that standpoint, most scholars consider qualitative methods inaccurate to establish or test a causal relationship so they utilize them, at best, as a complement to statistics. For instance they may use process tracing as a secondary technique, to shed light on a particular aspect of probabilistic models built on econometrics (Collier 2011; Seawright 2016). Even when they value the probatory contribution of case studies (Goertz and Mahoney 2012; Goertz and Starr 2003; Mahoney 2001) their search for sufficient and necessary conditions still requires the kind of cross-case comparison driven by a probabilistic logic of causation. For instance they may combine process tracing with comparative historical analysis and political economy (Falleti 2010; Hall 2012; Mahoney 2012).

4.3   The Problem of Case Selection

This handbook focuses more on small-N research design than most discussions on comparative methods in social sciences. For the reasons discussed already, such as the importance of context, we consider it crucial to think about how to conduct a research that looks carefully at a limited number of cases. The obvious example of the kind is the case study, but other methods such as ethnography also tend to focus only on a few cases, as they attempt to understand them more completely than might be possible through a large-N research design.

Once we decide to conduct a small-N comparison, a number of subsidiary questions arise about how this is to be done. The first is, which case or cases are to be considered? This question is addressed in some greater detail in the chapters on the comparative method and the most-similar and most-different systems designs, but here we point out that cases are only to some extent natural occurrences – cases have to be *constructed* by the researcher (Ragin and Becker 1992). Further, scholars and practitioners must decide whether their case is a case of some particular process or attribute. The same set of facts can be used in a variety of different ways to address different research questions.

A second dilemma raised by case selection is whether to use most-similar or most-different cases for comparison. The usual approach among social scientists has been to pick most-similar cases, and to use their similarity as a means of controlling extraneous variance. Then again, it is up to the researcher to make the case for similarity, and cases that are most-similar for some purposes could be most-different for others. Some scholars have argued in favor of utilizing most-different case designs (Przeworski and Teune 1970), with the argument being that if a relationship holds up across most-different cases then it is robust and can be seen almost as a law of political behavior.

Another issue which arises in case selection is the possibility of identifying a crucial case (Eckstein 1975), or a case that can be used to make a definitive test of a proposition. The design is, for example, to find a case in which the hypothesized relationship is least likely and determine if it is, in fact, supported. Some have argued that the possibilities of identifying these crucial cases are remote and consequently we should abandon that pursuit in favor of "pathway cases" (Gerring 2007) that can be used to elucidate the causal mechanisms involved in a theory rather than confirming or disconfirming that theory.

Finally, there is a question of linking within-case and cross-case analysis. Most case study methods focus on examining relationships and testing propositions within the single case. But for comparative analysis it is important to compare patterns across cases. Cross-case CPA allows us to understand the effects of context on a causal relationship, and to take into account other possible explanations for the findings in any one case. This cross-case comparison is central to Lijphart's comparative method, which he identified as one of the four fundamental research methods in the social sciences.

In a nutshell, each research design raises a specific question, which requires a specific set of methods to be aligned with a type of methodology according to the meaning of causality and causation. A general theory of causation describes regularities but it does not identify causal forces so it is not explanatory in a deterministic sense. A middle-range theory of causation describes a necessity, hence it explains how the qualitative nature of a social phenomenon varies according to the context, but its generalization is limited by contingency. Understanding the complementarities of these theories is paramount, if CPA scholars wish to preserve the coherence and consistency of their research design, especially when opting for

a multi-methods strategy. This applies both to dualists, who seek to produce explanations, and monists, who contend that only interpretations are in the realm of the social sciences. It also applies both to phenomenalists, who consider the only actually scientific knowledge is based on directly observable data, and transfactualists, who contend that this knowledge is also based on detectable though non-directly observable data.

5.   CONCLUSION: THE IMPORTANCE OF METHOD FOR COMPARATIVE POLICY RESEARCH AND PRACTICE

We shall conclude this introductory study by returning to some of the central contentions of the handbook. To characterize the current diversity of methods at the disposal of CPA scholars, it appears that the same biases prevail that exist in comparative politics, and in social sciences generally speaking. There is a persistent orthodoxy, favoring quantitative analysis and rejecting the possible contributions of qualitative techniques and multi-methods. Many high-impact journals these days are unfortunately filled with research articles that have followed a neo-positivist orthodoxy, hence failing to consider the full range of options and, perhaps more important, often failing to think about triangulation and the use of multi-methods as a means of gaining a more complete picture of the policy issue under scrutiny.

Yet at the same time it has become entrenched, there are more heretics providing alternatives to the received wisdom, whose heresies are supported by the existence of powerful qualitative methods, but perhaps even more by the increased recognition of the need to understand more about the process of policy making and the role of somewhat amorphous concepts – such as culture – on policy. Considering its multi-disciplinary nature, CPA is arguably more akin to multi-methods than any other social sciences areas. For one thing, context is important and bringing it in may effectively involve more qualitative understanding (Pollitt 2013). Moreover, CPA often involves a detailed examination of the policy process, which almost inherently requires qualitative methods such as congruence analysis and process tracing, small-N comparison, ethnography, etc.

As the content of this handbook indicates, we are ourselves very open-minded about which methods to use in CPA and above all we intend to make a modest but decisive contribution to the methodological debate. The book is organized in six parts. **Part I** is dedicated to the current methodological debate in social sciences that interests CPA. **Part II** draws on the contemporary trends in the methods and research agendas of some major analytical frameworks, related to comparison and causality in policy studies. **Part III** raises the problem of measurements and experiments in CPA. **Part IV** deals with mixed- and multi-methods in CPA research designs. **Part V** presents outstanding qualitative techniques more and more utilized in CPA. **Part VI** offers further reflections on the development of the field and the methodological debate thereof.

What we are advocating, above all, is a careful and conscious choice of the methods to be used, given the strength that triangulation of methods, and theory, can bring to research. On their own, all methods are not inherently good or bad, powerful or weak. What matters is under what conditions they are used and how well they are applied. The conscious choice of methods is important inasmuch as these have a decisive influence on our findings. Methods are not neutral instruments that will produce "correct" results, but they rather have their own biases and their own blind spots. For example, most quantitative methods assume an

additive relationship among variables, while QCA assumes a more interactive relationship, in which combinations of variables are associated with outcomes, and the absence of one or another produces no result, rather than merely a weaker result. Likewise, within-case study techniques look for particular decisions that produce the final outcome, rather than assuming that an outcome is a conjunction of variables. Both QCA and within-case methods are more deterministic than stochastic, hence they will provide a different type of understanding of the relationships among variables than will additive, statistical methods.

The methodological issues become more pressing when we attempt to explain differences across units, be they states, provinces, or whatever. As is discussed in several chapters of this handbook, case selection is crucial for conducting good comparative research, whether on public policy or any other topic. Cases picked simply out of habit or convenience are unlikely to provide the type of control over variance that is crucial for comparative analysis. Likewise, measurement is made more difficult by engaging in comparative analysis. Concepts may not "travel" well in comparative research (Sartori 1970), and that is even more of an intense problem for the measurement of those concepts, whether the measurement is qualitative or quantitative. A useful indicator in one context may be meaningless in another social system or culture, so the comparative researcher must pay particular attention to the meaning and context of the instruments used to measure.

The catalog of methods and techniques presented here should make it clear that there is a rich array of possible choices for the researcher. This handbook provides a good description and assessment of the methods available for CPA at this time. There are other methods that could have been included, but we do cover those most commonly in use for comparative analysis. Ultimately, this field is changing fast (see **Chapter 21** by Geva-May, Hoffman, and Muhleisen), so we need to be aware that those changes will affect the choice of methods. As already noted, the expanding use of experimental methods and the general development of qualitative techniques are enriching the methodological debate on CPA. The future for research appears extremely bright given the continuing development of techniques for understanding policy and the policy process. But, since choice can also be difficult, scholars must understand a range of alternatives in order to make the most appropriate choices for the problem at hand.

NOTES

1. Source: WOS database accessed on April 11, 2019.
2. N=79,471. Source: WOS database accessed on April 11, 2019.
3. The "critical" realist argument against Hume's "empirical realism" is that the world is an open system where regularities are more the exception than the rule, and our knowledge about the world cannot be reduced to atomistic events (Archer et al. 1998; Sayer 1992). Nevertheless we refer to "realism" instead of "critical realism", thus following the original statement of a "realist scientific endeavor" (Bhaskar 1978). The opposition of a "critical" realism to Hume's "empirical" realism is theoretically misleading, as it should apply to all but one methodologies. As a matter of fact scholars grounded in a reflexivist methodology owe a great deal to the critical theory which gave initially way to critical realism.

REFERENCES

Archer, M., Bhaskar, R., Collier, A., Lawson, T., and Norrie, A. (Eds.) (1998). *Critical Realism: Essential Readings*. London: Routledge.

Bardsley, N. (2010). *Experimental Economics: Rethinking the Rules*. Princeton, NJ: Princeton University Press.

Beach, D. and Pedersen, R. B. (2016). *Causal Case Study Methods: Foundations and Guidelines for Comparing, Matching and Tracing*. Ann Arbor, MI: University of Michigan Press.

Berg-Schlosser, D. (2013). *Process Tracing Methods: Foundations and Guidelines*. Ann Arbor, MI: University of Michigan Press.

Bhaskar, R. (1978). *A Realist Theory of Science*. London: Routledge.

Blatter, J. and Haverland, M. (2012). *Designing Case Studies: Explanatory Approaches in Small-N Research*. Basingstoke: Palgrave Macmillan.

Blatter, J. and Haverland, M. (2014). Case studies and (causal-) process tracing. In I. Engeli and C. Rothmayr Allison (Eds.), *Comparative Policy Studies: Conceptual and Methodological Challenges*. Basingstoke: Palgrave Macmillan, 59–84.

Brady, H. E. and Collier, D. (2010). *Rethinking Social Inquiry: Diverse Tools, Shared Standards*. Lanham, MD: Rowman & Littlefield.

Cairney, P. (2016). *The Politics of Evidence-Based Policy Making*. London: Palgrave Macmillan.

Carter, P. (2012). Policy as palimpsest. *Policy & Politics* 40: 423–43.

Castles, F. G. (2004). *The Future of the Welfare State: Crisis Myths and Crisis Realities*. Oxford: Oxford University Press.

Collier, D. (2011). Understanding process tracing. *Political Science and Politics* 44 (4): 823–30.

Dunlop, C. A. and Radaelli, C. M. (2013). Systematizing policy learning: from monolith to dimensions. *Political Studies* 61: 599–619.

Eckstein, H. (1975). Case studies and theory in political science. In F. I. Greenstein and N. W. Polsby (Eds.), *Handbook of Political Science*, vol. 7: *Political Science: Scope and Theory*. Reading, MA: Addison-Wesley, 79–133.

Engeli, I. and Rothmayr Allison, C. (Eds.) (2014). *Comparative Policy Studies: Conceptual and Methodological Challenges*. Basingstoke: Palgrave Macmillan.

Falleti, T. (2010). Infiltrating the state: the evolution of health care reforms in Brazil 1964–1988. In J. Mahoney and K. Thelen (Eds.), *Explaining Institutional Change: Ambiguity, Agency and Power*. Cambridge and New York: Cambridge University Press, 38–62.

Fischer, F. (2003). *Reframing Public Policy: Discursive Politics and Deliberative Practices*. New York: Oxford University Press.

Fontaine, G., Medrano, C., and Narváez, I. (2019). *The Politics of Public Accountability: Policy Design in Oil-Exporting Countries from Latin America*. Cham: Springer.

Gerring, J. (2007). Is there a (viable) crucial case method? *Comparative Political Studies* 40: 231–53.

Gerring, J. (2010). Mere description. *British Journal of Political Science* 42: 721–46.

Goertz, G. and Mahoney, J. (2012). *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences*. Princeton, NJ: Princeton University Press.

Goertz, G. and Starr, H. (2003). Introduction: necessary condition logics, research design, and theory. In G. Goertz and H. Starr (Eds.), *Necessary Conditions: Theory, Methodology, and Applications*. Lanham, MD: Rowman & Littlefield, 1–24.

Hall, P. (2012). Tracing the progress of process tracing. *European Political Science* 12: 20–30.

Hogwood, B. W. and Peters, B. G. (1987). *Policy Dynamics*. Brighton: Wheatsheaf.

Jackson, P. T. (2016). *The Conduct of Inquiry in International Relations: Philosophy of Science and Its Implications for the Study of World Politics* (2nd edition). London: Routledge.

Kellstedt, P. and Whitten, G. (2013). *The Fundamentals of Political Science Research*. New York: Cambridge University Press.

King, G., Keohane, R. O., and Verba, S. (1994). *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.

Mahoney, J. (2001). Path-dependent explanations of regime change: Central America in comparative perspective. *Studies in Comparative International Development* 36 (1): 111–41.

Mahoney, J. (2012). The logic of process tracing tests in the social sciences. *Sociological Methods & Research* 41 (4): 570–97.

March, J. and Olsen, J. (2006). The logic of appropriateness. In M. Moran, M. Rein, and R. E. Goodin (Eds.), *The Oxford Handbook of Public Policy*. New York: Oxford University Press, 689–708.

Pawson, R. (2006). *Evidence-Based Policy: A Realist Perspective*. London: Sage.

Peters, B. G. (2013). *Strategies for Comparative Research in Political Science*. Basingstoke: Palgrave Macmillan.

Peters, B. G. and Zittoun, P. (2016). *Contemporary Approaches to Public Policy*. London: Palgrave Macmillan.

Pollitt, C. (2013). *Context in Public Policy and Management: The Missing Link?* Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing.

Przeworski, A. and Teune, H. (1970). *The Logic of Comparative Social Inquiry*. New York: John Wiley.

Ragin, C. (2008). *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. Chicago, IL: University of Chicago Press.

Ragin, C. C. and Becker, H. S. (1992). *What is a Case? Exploring the Foundations of Social Inquiry*. Cambridge: Cambridge University Press.

Rihoux, B. and Grimm, H. (Eds.) (2006). *Innovative Comparative Methods for Policy Analysis*. New York: Springer.

Rose, R. (1993). *Lesson Drawing in Public Policy: A Guide to Learning Across Time and Space*. Chatham, NJ: Chatham House.

Sartori, G. (1970). Concept misformation in comparative research. *American Political Science Review* 64: 1033–53.

Sayer, A. (1992). *Method in Social Science: A Realist Approach* (2nd edition). London: Routledge.

Sayer, A. (2000). *Realism and Social Science*. London: Sage.

Scharpf, F. W. (1997). *Games Real Actors Play: Actor-Centered Institutionalism in Policy Research*. Boulder, CO: Westview Press.

Schneider, A. and Ingram, H. (1997). *Policy Design for Democracy*. Lawrence, KS: University Press of Kansas.

Seawright, J. (2016). *Multi-Method Social Science: Combining Quantitative and Qualitative Tools*. Cambridge: Cambridge University Press.

Shedish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Belmont, CA: Wadsworth.

Starke, P. (2006). The politics of welfare state retrenchment: a literature review. *Social Policy & Administration* 40: 104–20.

True, J. L., Jones, D., and Baumgartner, F. R. (2007). Punctuated-equilibrium theory: explaining stability and change in public policymaking. In P. A. Sabatier (Ed.), *Theories of the Policy Process*. Boulder, CO: Westview Press, 155–88.

Van Kersbergen, K. and Vis, B. (2014). *Comparative Welfare State Policies*. Cambridge: Cambridge University Press.

Webb, E. J., Campbell, D. T., Schwartz, R. D., and Sechrest, L. (1999). *Unobtrusive Measures*. Newbury Park, CA: Sage.

Weible, C. M. and Sabatier, P. A. (2017). *Theories of the Policy Process* (4th edition). Boulder, CO: Westview Press.

WOS (Web of Science) (2019). ISI Database. Accessed April 2019 at: https://clarivate.com.

APPENDIX

Table 1A.1   *Methods used for CPA by substantive areas (1945–2018)**

Area	Total (N=79,471)	Case study OR small-N (N=28,330)	Experiment (N=23,106)	Quantitative methods (big data OR econometrics OR time-series OR standard linear regression) (N=9,253)	Ethnography (N=2,489)	Focus groups (N=6,453)	Historical analysis (N=2,376)	Mixed methods OR multi-methods (N=3,599)	Q-methodology (N=2,745)	Comparative method OR most-similar/most-different systems design (N=310)	Process tracing (N=440)	QCA (N=370)
		Group 1	Group 1	Group 2	Group 3	Group 3	Group 3	Group 3	Group 3	Group 4	Group 4	Group 4
Environmental sciences and studies	0.19	0.28	0.13	0.18	0.04	0.11	0.17	0.13	0.16	0.13	0.07	0.19
Economics	0.12	0.08	0.15	0.38		0.22	0.17				0.07	
Public environment and administration	0.07	0.05	0.04	0.05	0.07		0.10	0.17		0.16	0.14	0.16
Political science	0.07	0.06	0.08	0.09	0.04		0.10	0.11	0.09	0.11	0.41	0.32
Education	0.04	0.06			0.11	0.09			0.16			0.05
Management	0.04	0.05	0.04									0.05
Geography	0.03	0.06		0.09	0.09		0.08		0.06			
Computer science	0.03		0.10									
International relations	0.01			0.03	0.08		0.06	0.05	0.06	0.05	0.23	0.12
Health policy	0.01					0.12		0.08		0.06	0.05	
Health care sciences	0.01					0.09		0.10		0.07		
Sociology	0.01				0.11		0.06		0.09		0.05	0.09
Anthropology	0.01				0.18							
Linguistics	0.01				0.04				0.12			
Communication	0.01								0.11			
History	0.01						0.05					
Other	0.33	0.35	0.46	0.18	0.24	0.36	0.21	0.37	0.14	0.42	0.00	0.00
Total	0.67	0.64	0.54	0.82	0.76	0.63	0.79	0.64	0.85	0.58	1.00	1.00

Note: *All key words associated with the descriptor "policy" in the query among "articles topics".

Source: Web of Science Database accessed April 22, 2019.