

# Exploring Pre-registration and Pre-analysis Plans for Qualitative Inference

Florian G. Kern<sup>†</sup>  
University of Essex

Kristian Skrede Gleditsch  
University of Essex

*July 25, 2017*

## Abstract

In recent years, the discipline of political science has experienced demands and moves toward greater research transparency. While in quantitative research increased transparency through replication has become a fairly accepted convention, in qualitative political science, the debate over whether and how to practice transparency is on-going. The practical tools and guidelines put forward to increase reliability of qualitative work so far emphasize a later stage of research, i.e. the phase of data analysis and sharing transcripts. To complement these suggestions, in this paper we explore how pre-registration and pre-analysis plans – as more recently introduced in experimental social science – can foster production transparency and analytic transparency in qualitative research. We argue that a “qualitative version” of pre-registration and PAPA can potentially incorporate a variety of approaches to inference. We discuss the general benefits and costs of such tools, and how they could inform qualitative research. Then, we provide a pre-registration template for such purposes.

## KEYWORDS:

transparency, replication, qualitative inference, pre-analysis plans, pre-registration

## *Acknowledgements:*

We thank the participants of the workshop “From Transparency to Replication in Qualitative Inference: Guidelines and Tools” (University of Essex, May 23-24, 2016) for comments and discussion. In particular, we are grateful to Jake Bowers, Edzia Carvalho, Louise Corti, Han Dorussen, Colin Elman, Alexandra Hartman, Nahomi Ichino, Sebastian Karcher, David Mainwaring, Jørgen Møller, Thom Oliver, Svend-Erik Skaaning and Kristi Winters for their feedback on an earlier version of this manuscript. Moreover, we are grateful to Heather Lanthorn, Rafael Piñeiro, and Fernando Rosenblatt for sharing their ideas on pre-registration for qualitative researchers. Finally, we thank the Pro-Vice-Chancellor (Research) at the University of Essex for providing funding for this research.

<sup>†</sup>Corresponding author, contact: [fkern@essex.ac.uk](mailto:fkern@essex.ac.uk)

# 1 Introduction

In recent years, the discipline of political science has experienced demands and moves toward greater research transparency (Lupia and Elman, 2014), with the goal to ultimately generate more replicable and reliable analyses. Leading journals now require authors to share their data and step-by-step analysis upon publication of articles, or even at the manuscript review stage. New principles of Data Access and Research Transparency (DA-RT) were partly incorporated in the ethics guide of the American Political Science Association (APSA) in 2012, emphasizing researchers' "ethical obligation to facilitate the evaluation of their evidence-based knowledge claims through *data access, production transparency, and analytic transparency* so that their work can be tested or replicated" (APSA, 2012, 9, added emphasis).<sup>1</sup>

On the one hand, for scholars employing quantitative approaches to social inquiry, i.e. "survey research, experiments, and mathematical and computerized models of dynamic processes" (Lupia and Alter, 2014, 54), data access and sharing for replication are fairly established conventions today. On the other hand, in qualitative political science, the debate over whether and how to practice transparency with regard to the sharing of research designs, data, and analyses, is on-going (Elman and Kapiszewski, 2014; QMMR, 2012, 2015). To increase transparency in qualitative research, political scientists have presented practical guides to sharing qualitative data (Elman and Kapiszewski, 2013), active citation (Moravcsik, 2010; Moravcsik, Elman and Kapiszewski, 2013), interview methods appendices (Bleich and Pekkanen, 2015), and introduced the *Qualitative Data Repository* (QDR) hosted at Syracuse University.

These practical tools and guidelines mostly support research transparency at a later stage of research, i.e. the phase of data analysis and sharing transcripts, field notes, etc. To complement these suggestions, in this paper we explore how pre-registration and pre-analysis plans

---

<sup>1</sup> For a critical discussion on transparency in qualitative research and the DA-RT principles see e.g. Isaac (2015), QMMR (2012), and QMMR (2015).

(PAPs) – usually employed e.g. in medical drug trials or by experimental social scientists – can foster production and analytic transparency in qualitative research. We argue that a “qualitative version” of pre-registration and PAPs can potentially incorporate a variety of approaches to inference. While we also acknowledge the emanating costs of these practices to the researcher, we show how pre-registration and PAPs can be beneficial both for hypothesis-testing and hypothesis-generating designs. They allow to increase the transparency and reliability of inference by explicating research design and methods prior to data collection and analysis, thereby addressing concerns that “[w]hat scholars do when conducting research, how they generate data, and how they make interpretations or draw inferences on the basis of those data, are rarely addressed at length in their published research” (Elman and Kapiszewski, 2014, 43).

In the following, we briefly summarize existing works on pre-registration and PAPs in the social science literature, and point to the benefits and costs of employing such devices. Then, we discuss how these tools can potentially be applied in qualitative research. We suggest that a new or any of the existing data repositories could host registrations and PAPs as well. For this purpose, we provide a preliminary template for pre-registration and PAPs in qualitative inference. Such template may serve as a baseline to contribute to the discussion of how applied pre-registration may work in qualitative inference (see also Piñeiro and Rosenblatt, 2016).

## **2 Pre-registration and pre-analysis plans**

Pre-registration and PAPs are fairly established in clinical trials (De Angelis et al., 2005), but have also recently been introduced in the social sciences, e.g. with the possibility to pre-register experimental research designs and analysis plans with the Evidence in Governance and Politics network (EGAP, see [www.egap.org](http://www.egap.org)). Only recently, the journal *Comparative Po-*

*litical Studies* has launched a call for special issue with a focus on transparency and pre-registration.<sup>2</sup> *Pre-registration* implies that researchers record research questions, design, and intended methods (and further information) of a study in an electronically searchable and accessible public database (Humphreys, Sanchez de la Sierra and van der Windt, 2013, 4). Usually, pre-registered designs are time-stamped and may be publicized once the researcher allows this (McKenzie, 2012). *Pre-analysis plans* can be part of pre-registration, and involve “a credibly fixed plan of how a researcher will collect and analyze data, which is submitted before a project begins” (Coffman and Niederle, 2015, 81). While both terms are at times used interchangeably in the literature, PAPs are somewhat less flexible (i.e. more binding) than a general pre-registration. Together, both tools can be used to pre-specify research questions, hypotheses, variables, data sources, sampling procedures, survey questions, data analysis procedures, or model specification (McKenzie, 2012; Olken, 2015).

Social scientists have discussed the benefits and costs of pre-registration and PAPs for research communities at length. Miguel et al. (2014) maintain that pre-registration can increase research transparency as a preceding step accompanying policies for open data (targeted at the reproduction of findings). More specifically, Findley, Humphreys and Weinstein (2012), Humphreys, Sanchez de la Sierra and van der Windt (2013), and Ozler (2015) argue that both tools tackle the problem of data and results “fishing” in experimental and observational studies. Beyond avoiding fishing, according to Coffman and Niederle (2015, 89), Gelman (2013, 40), and Olken (2015), generating rigorous PAPs may also help researchers to specify their designs in advance, and may trigger increased pre-testing. In addition, for researchers working with policy stakeholders, pre-specification of research designs and analyses can protect against conflicts of interest in case researchers find inconvenient results (Casey, Glennerster and Miguel, 2012, 1759). Olken (2015) summarizes the benefits of pre-registration and PAPs as follows:

---

<sup>2</sup> See <http://www.ipdutexas.org/cps-transparency-special-issue.html>

For readers, referees, editors, and policymakers, knowing that analysis was pre-specified offers reassurance that the result is not a choice among many plausible alternatives, which can increase confidence in results.

Others are more skeptical about the need and potential of these tools. Laitin (2013, 43) points out that – in contrast to clinical trials – the same protocol can rarely be applied to the same studies in the social sciences given varying contexts. Furthermore, too rigorous registration regimes may stifle the research process, since “[g]oing back and forth between exploration and testing is our disciplinary mode” (Laitin, 2013, 43). The fear here is that ‘tying one’s hands’ via pre-registration and PAPs may favor “robotic” hypothesis-testing designs and discriminate against exploratory studies and accidental findings, which produce important insights (see Gelman cf. 2013, 40; Anderson 2013). Moreover, because social inquiry is complex, writing a fully specified pre-analysis plan, taking into account all possible conditions and contingencies, is at least very costly and – in most cases – impossible (Olken, 2015, 70). Other concerns relate to practical issues of pre-registration – researchers may not want to share their designs early on with other scholars fearing their ideas may be copied (Coffman and Niederle, 2015, 90).

Then again, proponents of pre-registration and PAPs suggest flexible registries that distinguish “between pre-analysis and post-analysis hypotheses” to permit hypothesizing after the data is collected and analyzed (Humphreys, Sanchez de la Sierra and van der Windt 2013, 11; see also Casey, Glennerster and Miguel 2012). Miguel et al. (2014, 2) argue that pre-specification can allow a more credible differentiation between hypothesis-testing and hypothesis-generating designs, and thereby encourages exploratory work (see also Coffman and Niederle, 2015, 89). As Humphreys, Sanchez de la Sierra and van der Windt (2013, 2) emphasize, pre-registration and PAPs should be understood *both* as a commitment device as well as a tool for transparent communication of the research process.

### 3 Why pre-register qualitative designs?

While some authors formulate general calls for pre-registration in political science (e.g. Findley, Humphreys and Weinstein, 2012), few studies address pre-registration and PAPs for qualitative research extensively. A notable exception are Piñeiro and Rosenblatt (2016, 788), who share our idea that qualitative inference “does not entail that everything be learned or done in the field or through the compilation of secondary sources”, but that many facets of qualitative designs can be determined and communicated before the back and forth between theory, empirics, and interpretation. They suggest and thoroughly discuss the use of pre-analysis plans. Further, Olken (2015, 69) discusses the combination of qualitative evidence and experimental trials. McDermott (2014, 70) mentions the risk of revealing information about interview sources if transparency was favored over sensitivity and identity issues in qualitative research. Humphreys, Sanchez de la Sierra and van der Windt (2013) highlight that pre-registration could emphasize the difference between exploration and tests of hypotheses. Miguel et al. (2014) generally call for further exploration of pre-registration for qualitative data. Monogan (2013, 35) proposes that qualitative field research can benefit from pre-registration as well.

Some of the arguments in favor of pre-registration and PAPs outlined above also apply to qualitative inference, especially when the latter is aimed at empirical application and plausibility probes. While straightforward replication as a tool to ensure transparency is more difficult for many qualitative than for quantitative approaches, a flexible implementation of pre-registration and PAPs might increase the reliability of qualitative inference, as it emphasizes transparency at an earlier stage of the research process. A transparent archiving of research questions, expected data generating processes, and intended designs could help in providing insights into the procedure of qualitative inference. Together with presenting the post-registration analysis, scholars could illustrate their inferential process, and credi-

bly point to surprising findings and hypotheses generated through the analysis of data or experiences in the field.

To name but a few examples on how pre-registration and PAPs could be utilized in qualitative research:

- If employing *case studies*, researchers could become more transparent about whether their designs are aimed at being hypothesis-generating or hypothesis-testing (i.e. plausibility probes), which variables were used for the selection of cases and why, and which methodologies or sources will be used to tackle a research question.
- If applying *process-tracing*, scholars could make their priors about causal processes more explicit by pre-registering their theoretical expectations.
- If conducting in-depth *qualitative interviews* or *focus groups*, researchers can share their questionnaires or guidelines, or how they plan to produce interview samples, and thereby be more transparent about the overall research process.
- If using *active citation* or *Computer Assisted Qualitative Data Analysis (CAQDAS)* packages to analyze their sources, scholars can share their analytical categories in advance.
- If conducting *ethnographical fieldwork*, investigators can (as suggested by Monogan, 2014) communicate their target populations and locations, as well as their observational or participatory approach.
- If employing *mixed methods designs*, researchers can credibly claim that qualitative evidence informed a later quantitative analysis, or vice versa.

Qualitative researchers with empirical research interests could use pre-registration while staying flexible about their analysis, since they could emphasize which findings correspond to pre-registered designs, and which insights are obtained by post-data collection or post-analysis hypothesizing. Pre-registered hypotheses could be reformulated after a learning process through data analysis has taken place. Furthermore, readers could evaluate the in-

ternal validity of the concepts used in a qualitative study if they have access to both the pre-registered design and the following analysis. Used this way, pre-registration and PAPs in qualitative inference closely correspond to the ideal research “disciplinary mode” of political science, outlined by Laitin (2013) as oscillating between theorizing, exploration, and testing, while adhering to a transparent approach to inference.

To be sure, issues of the sensitivity of research design and data sources are valid concerns for qualitative researchers. Pre-registration may be able to incorporate some of these matters. On the one hand, time stamped pre-registration may allow scholars to archive their designs and only open them to the public when analysis or manuscript publication are imminent. And even if transcripts and lists of participants contain sensitive information about subjects and cannot be shared following the study, one might argue that – all else equal – prior transparency based on a maximum number of parameters of the study increases reliability of qualitative inference even without availability of verbatim transcripts.

## **4 Pre-registration Template**

Below, we list possible elements of a pre-registration form. Here, we do not have to re-invent the wheel. Several authors have put forward checklists and templates for pre-registration of experimental and observational studies (EGAP 2016; Findley, Humphreys and Weinstein 2012; Humphreys, Sanchez de la Sierra and van der Windt 2013, 12ff.; McKenzie 2012; Olken 2015, 65). We use these as a basis for our template, but alter certain elements to incorporate qualitative research designs. The registration form would be time stamped and receive a unique identification code. If researchers were to adjust their research plans, they would be allowed to annotate their initial registration. One might also think of asking researchers every six months to check their registration materials and update them if necessary. We suggest the following template could provide a frame for pre-registration and PAPs



in qualitative research.

<b>A. General information</b>	
A.1 Project title:	
A.2.1 Authors:	A.2.2 Affiliation:
A.3 Contact details:	
A.4 Abstract and research question(s):	
A.5.1 Project start date:	A.5.2 Estimated project end date:
A.6.1 Should this study be gated?	A.6.2 If yes, until when?
A.6.3 Has this research received approval from an Institutional Review Board (IRB) or ethical board?	
A.6.4 If yes, please name date and institution of approval:	
<b>B. Research design specifics</b>	
B.1 Please describe the data generating process of the study:	
B.1.1 Please describe the unit of analysis of the study:	
B.2 Is this project hypothesis-testing or hypothesis-generating (or both)?	
B.3.1.1 If hypothesis-testing, please name and describe the hypotheses below:	
B.3.1.2 If hypothesis-testing, please describe the methodology you employ for the test:	
B.3.2.1 If hypothesis-generating, please name and describe possible outcome hypotheses:	
B.3.2.2 If hypothesis-generating, please describe the methodology you employ to generate hypotheses:	
B.3.3 Please define and describe the outcome variable or concept in your hypotheses:	
B.3.4 Please define and describe the independent variables or concepts in your hypotheses:	
B.4.1 Does this project involve fieldwork?	B.4.1.1: If yes, where?
B.5 Does this project involve archival work?	B.5.1: If yes, where?
B.5.2 If yes, what are the archival sources you will use?	
B.6 Does this project involve interviews or focus groups?	
B.6.1 If yes, please describe the target population:	B.6.2 If yes, what is the targeted sample size?
B.6.3 If yes, please describe the sampling strategy?	
B.6.4 If yes, please attach a sample questionnaire or sample guidelines:	
B.6.5 If yes, please attach a copy of your subjects' informed consent form	
B.7 Does this project involve causal process tracing?	

B.7.1 If yes, please describe possible outcome causal processes:
B.8 Does this project involve case studies?
B.8.1 If yes, please describe the selection process and criteria for your cases:
<b>C. Analysis</b>
C.1 Please describe the analytical tools you use for this project:
C.2 Does this project involve computer assisted qualitative data analysis?
C.2.1 If yes, what software package do you use?
C.2.3 If yes, what are the analytical categories and codes used for the analysis?
C.3 Does this project involve active citation?
C.3.1 If yes, give a specific example for how you employ active citation
C.3.2 If yes, is this project registered with the Qualitative Data Repository?
C.4 Does this project involve sensitive data you cannot share?
C.4.1 If yes, please describe the nature of this data and why you cannot share it:
C.5 Do you use any other kind of methodology in your study? (quantitative data, experiments, etc.)
C.5.1 If yes, please specify.
<b>D. Logistics</b>
D.1 Please define stages in the research process at which you would like to update your registered design:
D.2 If applicable, please name the policy organizations collaborating with you on the project:
D.3 If applicable, please name the source of funding for your project:
D.3.1 If applicable, please detail the research budget of your project:
D.4 Please indicate the type and number of outputs and publications planned for this project:
D.5 Please provide any additional comments on the registration of your research project:
D.6 Please provide any other supporting documents:
<b>CERTIFICATION</b>
<i>I hereby register my research project and supplementary materials. I confirm that I own the rights to release these materials into the public domain.</i>

## 5 Conclusion

In this paper, we explore how pre-registration and pre-analysis plans can foster production transparency and analytic transparency in qualitative research. We argue that a “qualitative version” of pre-registration and PAPs can potentially incorporate a variety of approaches to inference. We discuss the general benefits and costs of such tools, and how they could inform qualitative research. Then, we provide a pre-registration template for such purposes.

Beyond the guidelines we provide, it will be important to pilot the template in order to test its fit for various approaches to qualitative inference, which we plan to do in the future. We also encourage other researchers, both qualitative and quantitative, to join the discussion, criticize and ameliorate our template. Hoping to trigger a debate with this article, we are fully aware that that pre-registration and this template will appeal most to researchers who already hold an epistemological and methodological stance grounded in empirical research. To be sure, the use of pre-registration and PAPs by itself does not guarantee that inference will become more transparent or better at every step of the research process. Rather, a qualitative use of pre-registration as a tool for commitment *and* communication may benefit the transparency of qualitative research by showing the procedure of qualitative inference, and may thereby allow more structured, constructive criticism.

## References

- Anderson, Richard G. 2013. "Registration and Replication: A Comment." *Political Analysis* 21(1):38–39.
- APSA. 2012. *A Guide to Professional in Political Science*. Washington, D.C.: American Political Science Association.
- Bleich, Erik and Robert J. Pekkanen. 2015. "Data Access, Research Transparency, and Interviews: The Interview Methods Appendix." *Newsletter of the American Political Association Organized Section for Qualitative and Multi-Method Research* 13(1, Spring).
- Casey, Katherine, Rachel Glennerster and Edward Miguel. 2012. "Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan." *The Quarterly Journal of Economics* 127(4):1755–1812.
- Coffman, Lucas C. and Muriel Niederle. 2015. "Pre-Analysis Plans Have Limited Upside, Especially Where Replications Are Feasible." *The Journal of Economic Perspectives* 29(3):81–97.
- De Angelis, Catherine D, Jeffrey M Drazen, Frank A Frizelle, Charlotte Haug, John Hoey, Richard Horton, Sheldon Kotzin, Christine Laine, Ana Marusic, A John PM Overbeke et al. 2005. "Is this clinical trial fully registered? – A statement from the International Committee of Medical Journal Editors." *New England Journal of Medicine* 352(23):2436–2438.
- EGAP, Evidence in Governance and Politics Network. 2016. "Create Registration' Form." Online.  
**URL:** <http://egap.org/node/add/registration>
- Elman, Colin and Diana Kapiszewski. 2013. *A Guide to Sharing Qualitative Data*. Qualitative Data Repository (QDR), Center for Qualitative and Multi Method Inquiry (CQMI), Syracuse University. Version 1.3.
- Elman, Colin and Diana Kapiszewski. 2014. "Data Access and Research Transparency in the Qualitative Tradition." *PS: Political Science & Politics* 47(1):43–47.
- Findley, Michael, Macartan Humphreys and Jeremy Weinstein. 2012. "Draft proposal for a Pilot Registry for Political Science (PREPS)." *EGAP VIII Steering Group* October 3, 2012.
- Gelman, Andrew. 2013. "Preregistration of Studies and Mock Reports." *Political Analysis* 21(1):40–41.
- Humphreys, Macartan, Raul Sanchez de la Sierra and Peter van der Windt. 2013. "Fishing, Commitment, and Communication: A Proposal for Comprehensive Nonbinding Research Registration." *Political Analysis* 21(1):1–20.

- Isaac, Jeffrey C. 2015. "For a more *public* political science." *Perspectives on Politics* 13(2):269–283.
- Laitin, David D. 2013. "Fisheries Management." *Political Analysis* 21(1):42–47.
- Lupia, Arthur and Colin Elman. 2014. "Openness in political science: Data access and research transparency." *PS: Political Science & Politics* 47(1):19–42.
- Lupia, Arthur and George Alter. 2014. "Data Access and Research Transparency in the Quantitative Tradition." *PS: Political Science & Politics* 47(1):54–59.
- McDermott, Rose. 2014. "Research Transparency and Data Archiving for Experiments." *PS: Political Science & Politics* 47:67–71.
- McKenzie, David. 2012. "A pre-analysis plan checklist." The World Bank. Development Impact Blog.  
**URL:** <http://blogs.worldbank.org/impactevaluations/a-pre-analysis-plan-checklist>
- Miguel, Edward, Colin Camerer, Katherine Casey, Joshua Cohen, Kevin M Esterling, Alan Gerber, Rachel Glennerster, Don P Green, Macartan Humphreys, Guido Imbens et al. 2014. "Promoting transparency in social science research." *Science* 343(6166):30–31.
- Monogan, James E. 2013. "A Case for Registering Studies of Political Outcomes: An Application in the 2010 House Elections." *Political Analysis* 21(1):21–37.  
**URL:** <http://pan.oxfordjournals.org/content/21/1/21.abstract>
- Monogan, James E. III. 2014. "Political Analysis Using R."
- Moravcsik, Andrew. 2010. "Active Citation: A Precondition for Replicable Qualitative Research." *PS: Political Science & Politics* 43(01):29–35.
- Moravcsik, Andrew, Colin Elman and Diana Kapiszewski. 2013. *A Guide to Active Citation*. Qualitative Data Repository (QDR), Center for Qualitative and Multi Method Inquiry (CQMI), Syracuse University. Version 1.8.
- Olken, Benjamin A. 2015. "Promises and Perils of Pre-Analysis Plans." *The Journal of Economic Perspectives* 29(3):61–80.
- Ozler, Berk. 2015. "Preregistration of studies to avoid fishing and allow transparent discovery." The World Bank. Development Impact Blog.  
**URL:** <http://blogs.worldbank.org/impactevaluations/preregistration-studies-avoid-fishing-and-allow-transparent-discovery>
- Piñeiro, Rafael and Fernando Rosenblatt. 2016. "Pre-Analysis Plans for Qualitative Research." *Revista de Ciencia Política* 36(3).

QMMR. 2012. "Symposium: Research Openness in the Digital Age." *Newsletter of the American Political Association Organized Section for Qualitative and Multi-Method Research* 10(1, Spring):28–38.

QMMR. 2015. "Symposium: Transparency in Qualitative and Multi-Method Research." *Newsletter of the American Political Association Organized Section for Qualitative and Multi-Method Research* 13(1, Spring).