

Discussion document:

A Proposal for a Political Science Registry

Draft: August 2015

**Composed by a joint committee of APSA's Organized Sections for *Political Methodology*,
Qualitative and Multi-Method Research, and *Experimental Research*¹**

Note: This draft is for discussion purposes only. The proposal it contains has not been endorsed by the leadership of the three sections that appointed this committee or by their broader memberships. We welcome comments from members of these sections as well as the broader research community.

Executive Summary

We propose the establishment of a registry for political science research, the *Political Science Design Archive* (PSDA). The registry would allow researchers to report a study design and, if possible, a specific plan by which that study will be implemented. The registry would also allow for archiving of results of the plans, as well as results that do not appear in publication form elsewhere. The data might be experimental or observational, quantitative or qualitative, contemporary or historical. After research is complete, the registry would contain an archive of the results of the research, in brief form.

The first benefit of a PSDA would be to help clarify the distinction between research that is more exploratory and research that is more confirmatory (eg testing) so that the scholarly community can evaluate research findings more easily. For confirmatory research in particular, the following benefits may also be counted in favor of the proposal. (1) Transparency in the research process will be enhanced. (2) A venue will be created for the publication of findings that do not find their way into existing journals or books, thus addressing the “file-drawer” problem (engendered by publication biases in the discipline). (3) Based on these findings, meta-analyses may be facilitated, enhancing knowledge cumulation. (4) Replication will be facilitated.

We recognize the many challenges faced by this initiative, which is considerably more far-reaching than extant registries. To that end, we solicit input so that we can shape and refine this proposal to best serve the political science community.

¹ These three sections are represented by Jake Bowers and Jonathan Nagler (Section 10: Political Methodology), John Gerring and Alan Jacobs (Section 37: Qualitative and Multi-Method Research), and Don Green and Macartan Humphreys (Section 42: Experimental Research).

1 Introduction

Many research projects that are undertaken do not result in published findings, and many analyses that are implemented in any given research project do not get reported. Selective or haphazard reporting can give rise to different types of biases of interpretation and knowledge cumulation. Biases arise in particular if the decision to report or publish is a function of research findings.

A research registry --- which, in its most basic form, is simply a searchable archive of research designs --- might help address this problem in at least two ways. First, by archiving study designs, a registry generates a database of research projects that is not conditioned by findings. This increases the discoverability of research, and can provide a tool to better place individual studies in context. Second, if the registry also contains a record of researcher intentions *prior to analysis* a registry can help readers distinguish between exploratory and confirmatory analyses. The first of these activities (“pre-registration” archiving) consists of gathering basic data on research projects; the latter activity (“pre-analysis plan” archiving) requires much more detailed information on analysis strategies be provided.

These benefits have led to a growth in the adoption of registration practices. Registration has now become standard practice for medical trials, and since 2005 major medical journals have required registration of all randomized controlled trials (RCTs) as a precondition for publication. There has also been growing interest in registration in political science and other social sciences. The American Economics Association has established a [trial registry](#) (open only to RCTs), and the International Initiative for Impact Evaluation (i3e) has established the Registry for International Development Impact Evaluations ([RIDIE](#)) (open to experimental and observational studies assessing the impact of development programs). *Political Analysis* now encourages preregistration in its instructions to authors and many political science journals have become signatories to the [TOP guidelines](#), which commits them to reviewing and adopting a position on registration within the coming year. There has also been an increase in registration practices among researchers, with nearly 200 projects now registered on the [Experiments in Governance and Politics](#) website.

While few argue against the principle of analytic transparency, there are concerns around the idea of a registry. The key concerns are that a registry, if badly conceived, could stifle innovation, inappropriately constrain authors, exacerbate inequalities, introduce cumbersome obstacles to research, install unaccountable gatekeepers, create risks of intellectual property theft, and distract from more urgent priorities. There are also more mundane concerns around the logistics, staffing, and funding of a registry.

Finally, there are more fundamental concerns that registration may not be relevant for the vast majority of political science research, most of which is observational, and much of which is based on historical or qualitative data. As we argue below however, the central issue has to do with

confirmation versus exploration and not the mode of observation; for example in our review of 92 articles conducting substantive qualitative empirical analyses, 40 --- comprising 43% --- explicitly claimed to have used the qualitative evidence in a confirmatory fashion. The challenge for these researchers in distinguishing the testing from the induction is similar to that facing quantitative scholars. In light of this concern, we see the need to expand thinking around registration --- including its potential usefulness for some forms of historical and qualitative research --- as a critical challenge.

In January 2015, the presidents of three APSA methodology sections (Political Methodology, Qualitative and Multi-Method Research, and Experimental Political Science) jointly appointed an inter-section committee to explore the options. The deliberations of the group form the basis of this discussion document. These deliberations build on broader consultations including with political science journal editors (at EGAP 8 meetings in Stanford 2012), in the pages of *Political Analysis* (vol 21), on various blogs, and at a daylong symposium hosted by IQMR at the Maxwell School at Syracuse University. The document is to be presented to the presidents and councils of the sections, shared with other leading sections for comment, and available for general discussion at the 2015 APSR meetings in San Francisco. [This document is hosted on Google Docs and is open for general commentary.](#)

In summary, the proposals under consideration by the inter-sectional committee, elaborated in Section 5 below, are as follows:

1. An electronic registry will be established as a project of the three APSA methods sections.
2. This registry will be guided by a set of core principles: transparency, neutrality, navigability, complementarity, automation, and permanence. It aims first and foremost to provide a *facility*: first, to researchers seeking a place to archive designs in a central and credible repository; second, to journals seeking to share anonymized or other designs with reviewers or to confirm authors' registration claims; and third, to other researchers seeking to assess what research has been undertaken on different topics. The goal is not to evaluate research, constrain authors, or influence journal policies.
3. The registry should be a full-service, staffed facility that provides minimal supervision and, on request, confirmation of registration (a technical task defined below). It should allow users to post and amend designs, keep designs public or private for some period of time, and let authors share anonymous timestamped designs with editors and reviewers. It should provide a facility for authors to receive comments on registered designs and to connect their designs to summary reports, working papers, and data hosted on other repositories.
4. The registry should be governed by a non-academic director who answers to a board with members representing the three sections and a seat for an external stakeholder (such as a representative of other sections, of the APSA Council, or of an external research body such as the SSRC.)

In the remainder of this document we describe briefly what a registry is, or can be; the arguments in favor or against various functions; the current options available for registration in political science; and the outlines of our proposal with reflections on institutional structure.

2 What is a registry?

In its most basic form a registry is a lockbox for storing descriptions of research designs *whether or not the resulting research is published*. It has advantages when it is a *focal* lockbox, so that people know where to go to find out about research projects on a given topic. It has additional advantages when the lockbox timestamps documents, as this allows users to distinguish, if they wish, between research strategies worked out prior to analysis and those generated during the course of the analysis.

Beyond this simple lockbox function, we can identify nine other functions (described in Table 1 below) that a registry might have. Four **basic functions** are the ability to provide individual project registrations with unique *identifiers* so that these can be noted in research products or cited as objects; the ability to allow users to update designs and add related documentation in a way that preserves the history of changes; the ability to provide minimal supervision to ensure that registered projects record basic common information; the ability to be searchable, and in particular to allow registered documents to be searchable across registries.

More elaborate functions could allow users to control which information becomes public at what point in time; allow researchers to give some readers (such as editors or reviewers), but not others, access to registered documents; or connect registries to other documents such as published research as it becomes available, unpublished research--such as short accounts of findings whether or not these eventually get published --- or to data repositories, such as Dataverse, that store data and code.

Two other possible functions include one aimed at authors and one at journals.

For authors, a registry could provide a platform to host, and perhaps even moderate, non-anonymous private or public **comments** that provide opportunities for researchers to receive feedback even prior to research implementation.

For journals, a registry could provide **confirmation** that reported analyses are consistent with registered analyses. Note that in the table we distinguish between two forms of technical confirmation of registration:

- **COD**: basic confirmation of design registration, “COD” confirms that a general design corresponding to a research product was registered; this requires only minimal supervision.
- **CAP**: confirmation of analysis plans (“CAP”) confirms that analysis plans correspond to actual analyses with all modifications marked up (leaving to research consumers the task of assessing the quality of analyses and the justification for any modifications).

Table 1: Functions of a Registry

| | |
|--|--|
| Minimal: | |
| 1 A lock box | The most basic function of a registry is to provide a facility to deposit and store a publically accessible and timestamped collection of documents. This basic function can also be achieved using platforms such as dataverse or the Open Science Framework. |
| Core: | |
| 2 Provide Identifiers | Registered designs should have a unique identifier associated with them. If they have a <i>DOI</i> then they also become citable objects which facilitates referencing both by original researchers but also by others that may want to reference the design. |
| 3 Versioning | Researchers should be able to post revisions and updates, with time-stamping and versioning control. |
| 4 Supervised: Minimal Confirmation of Design registration [COD] | Supervision means that minimal data is required for registration and that the registry only contains minimally complete design descriptions. <i>Confirmation</i> of design registration (COD) does not imply any claim about the quality of a research design. |
| 5 Searchability | It could be possible to search a registry by keyword and also cross-search across registries. |
| Additional: | |
| 6 Control public access. | A registry could provide an option to researchers to deposit documents that are only available to the public after some specified period. |
| 7 Access anonymized versions | To facilitate reviewing, authors should be able to control who gets to see what. In particular authors should have access to a shareable link, possibly password protected, to a possibly anonymized version of the registry, even if the registry is gated. |
| 8 Connect designs to studies and data | A registry could link from a registration to subsequent working papers, research notes, papers, and data. |
| 9 Allow commenting | Researchers could benefit from an option to allow peer commenting on specific versions of a design. |
| 10 Confirmation of Analysis pre-registration [CAP] | Journals could benefit from outsourcing to registries the work of confirming whether a study pre-registered an analysis <i>plan</i> [CAP]. |

3 Arguments for and against registration

There are multiple arguments for and against the proposed registry. These different arguments relate to different functions of a registry. Some issues are more fundamental than others, and

some relate to issues of implementation rather than to the core question of whether a political registry ought to be formed.

3.1 Arguments supporting registration

We highlight three families of argument that support the principle of research registration.

Registration Helps Counter Publication Bias. The simplest argument for a registry is that it provides a location to record the existence of research projects whether or not they are published. This allows research consumers interested in a particular topic to assess research that is implemented whether or not it is ultimately published. This benefit relies on **Function 1** (lockbox) of a registry but also presupposes that a registry has reasonable focality and that it maintains a significant share of research undertaken on a topic. Note that for this purpose there is no requirement that a registry contains designs prior to implementation. To be useful for this purpose a registry should also be searchable (**Function 5**) and ideally include basic information about any unpublished results (**Function 8**).²

There is now substantial evidence of publication bias in political science. Some recently striking evidence from Franco et al (2014) found that, among experiments implemented via the TESS platform, studies with null results were significantly less likely to be published or even to be written up.

Registration Helps Counter Data Fishing. A distinct concern that motivates *pre*-registration in particular is the possibility that researchers engage knowingly or unknowingly in a type of informal data-mining. Thus researchers may engage in exploratory analysis of various forms but later present these as tests of hypotheses; or they may undertake multiple analyses, for example of different outcome measures, but focus write-ups on the significant or striking findings. There are sometimes structural incentives to adjust specifications in light of findings, for example in response to reviewer requests which are generally made after core results are known.

The scope for fishing of this kind is large when researchers have discretion to examine (but not report) many variations of a single analysis (see Humphreys et al 2013). There is little non-anecdotal evidence of data fishing. However, analyses of published statistics in political science journals (Gerber and Malhotra 2006) suggest a very strong tendency for published results to lie just below critical values such as $p = 0.05$ (and relatively few just above such thresholds), though it is difficult to know whether these patterns reflect analysis bias or publication bias. In one study, Gerber et al (2010) suggest that there is “evidence that papers systematically employ one-sided hypothesis tests in response to failure to meet the more demanding critical values associated with two-tailed tests.” Evidence of broad failure to replicate

² In the absence of information about results, a presumption would likely be that the research failed to confirm the stated hypothesis -- suggesting a null result which could inform future research.

published studies is also consistent with fishing, though it may also reflect fundamental instability (see, for example, results from the [Reproducibility Project](#)).

A registry can limit damage from data fishing by providing transparency in the research process. Registering a study does not require an obligation to adhere to those designs; it simply generates a record of intentions. As a communication tool, registration helps research audiences distinguish confirmatory and exploratory analyses, allowing researchers to credibly signal when they are undertaking the former. Yet it in no way inhibits authors of registered studies from conducting and reporting analyses that occur to them after they have seen the data. Pre-registration may be followed by “soaking and poking.” The point is that readers of the study will have a way to ascertain which elements were tested, based on an *ex ante* (registered) plan of data collection and analysis, and which elements were developed in dialog with the data.

Again Function 1 (lockbox) is most important for this benefit. Function 3 (versioning) can help limit the extent to which registration imposes constraints. For example, with sophisticated versioning, an author might be able to indicate the timing of a change to a design decision so that readers can assess whether those changes introduce concerns around inference or not.³ To have these benefits a registry would likely need to be able to store relatively detailed analysis plans and not simply general descriptions of research designs. And of course, registration should take place prior to analysis (and indicate more precisely when in the research process it was conducted; more on this below). Functions 6, 7, and 10 could help ensure that pre-registered analyses are available to editors and reviewers and that readers can receive summary assessments of whether research as reported, along with any modifications, is consistent with registered intentions.

Other Synergistic Benefits. Other benefits derive from other possible functions of registries, such as providing a forum for authors to receive comments on research designs prior to implementation, or for linking designs to subsequent research and data. Comments prior to implementation are often most useful since adjustments may be made in time to respond to important concerns. In the case of research that is implemented in partnerships, early posting of analysis details can help ensure that all parties have a common understanding of analysis plans and can shield researchers from pressures to adjust analyses *ex post*. Sometimes, pressures arise from funders, who may have strong views about what sort of finding is desirable or plausible. Registration of a detailed pre-analysis plan may help to get funders on board with a credible plan for testing, and will certainly make it more difficult to repress results. Finally registration may have benefits by drawing public attention to the planning and motivating of research by *ex ante* considerations related to substance and theory and to temper and contextualize novel findings.

INVITATION TO READERS: ADD MORE ARGUMENTS IN FAVOR IN COMMENTS →

³ For example a researcher might implement a two-wave study but add additional hypotheses after Wave 1 to be tested with Wave 2 data.

* * *

Many of these benefits for registration can be motivated by scholarly or ethical concerns. The scholarly argument is that these benefits advance learning and render research more credible. The ethical arguments are that researchers that have engaged human subjects (or funders, or colleagues, or others) to generate knowledge have an obligation to ensure that that research be reported without bias, no matter what the results of the research.

3.2 Arguments against

We have encountered at least ten arguments against the idea of registration.

1 Registration kills innovation. A common concern raised is that registration will kill innovation in two ways. First, one might worry that researchers analyzing data will refrain from exploratory work for fear that this looks like a deviation from their analysis plan. Second, researchers may not register research that is truly exploratory in nature and this work will ultimately be devalued relative to confirmatory research.

Both concerns have face validity. However, the aim of registration is to clarify when research is exploratory and when it is confirmatory, not to privilege one form of research over another.

It is possible, of course, that an unintended side effect of registration might be the appearance of the labeling of some forms of inquiry as more rigorous than others. We view this as a concern not about a registry per se but about current research norms. At present, theory testing is generally treated (at least in public venues) as a more valued activity than exploration and discovery. Nonetheless, political scientists routinely engage in exploratory research, and privately recognize its critical role in advancing research agendas. The problem, in our view, is that scholars often hesitate to label their own work as exploratory when placing it before the public. Registration could help to shift norms in a fruitful direction. By reducing the scope for credibly claiming to have tested a hypothesis, registration may help limit the language of theory-testing to mature stages of research programs, leaving space for alternative descriptions of research activities at other stages.

We also hope that this discussion about research registration helps us to conceptualize useful ways in which exploratory research could be registered. Most simply, researchers engaged in exploratory work could register that work in advance and could indicate the exploratory nature of the analysis. This would clarify the purpose of the work and limit scope for claiming ex post that the registered research was designed to do theory testing. It would also create a record of what exploratory work has been done and when it has been fruitful. A second approach is to register formal exploratory protocols, such as those used in machine learning or formal data mining.

2 Registration produces damaging inflexibility: A related concern is that registration will force researchers to implement registered analysis plans even though, once they see their data, they realise that these plans are not optimal. On this, Tucker writes:

“giving people a grade for faithfully sticking to a registry could lead to people conducting sub-optimal research — and stifle creativity — out of fear that it will cost them their “A” registry-faithfulness grade.”

<http://blog.oup.com/2014/09/pro-con-research-preregistration/#sthash.1FpPe27T.dpuf>

This concern is, again, not about a registry *per se*, but rather about how researchers, journals, and reviewers might perceive and make use of this new tool. In pursuing a registry for political science we would want to ensure that any systems established to *recognize* registration do not penalize (clearly signaled) deviations from registration plans. To illustrate, the Open Science Framework registration badge scheme awards badges to researchers even when they deviate from registered plans, *as long as those deviations are transparently documented*. In those cases the badges signal that there have been deviations with a ‘TC’ (transparent change) designation. It is up to the political science community to decide whether the ‘TC’ moniker is a dreaded asterisk or simply a suggestion to read the changes in more detail.

3 Registration creates an unaccountable gatekeeper. A third concern is that registration could establish unaccountable gatekeepers. This concern pertains especially to Function 10 of a registry, the confirmation of research designs. For example, Josh Tucker writes:

“I want to prevent some arbitrary third party from being given the authority to “grade” researchers on how well they stuck to their original design and then to be able to report that grade publically, as opposed to simply allowing readers to make up their own mind in this regard.”

<http://blog.oup.com/2014/09/pro-con-research-preregistration/>

This concern is perhaps best addressed by ensuring that any confirmation process is technical in nature, i.e., that it is a simple check on transparency claims. Confirmation would consist strictly of examining whether any reported analyses that authors claimed by to have been registered were in fact registered. In practice, ensuring no value judgements in the determination of fidelity (with transparent deviations) may be difficult. For instance, judgment may be involved in determining when small changes to analysis code or to the wording of hypotheses count as substantively meaningful deviations. The approach we favor for Confirmation of Analysis Plans (CAP) is to place strong demands on researchers to document transparent changes to code so that there are no substantive judgement calls for registry staff and to validate analysis consistency only when a research product is accompanied by a working replication archive. By providing both the pre-analysis plan code, and final code, and final code in annotated form showing changes -- deviations from pre-analysis plan are readily available for readers. Thus CAP may signify

stringent adherence to the original plan, or weak adherence - but the level of adherence will be transparent.

Implicitly what CAP means is a statement of the form:

"The author's factual statements about pre-registration are correct, and no important details have been omitted. We leave to readers the job of assessing whether any deviations from the original plan should be judged important."

4(a) Registration Will Not Stop Fraud. Some note that, if the purpose of registration is to prevent researchers from cheating in their analysis then it is not sufficient for the task. Unscrupulous researchers could always implement their analyses first, then register, claiming that the research had not yet been implemented. Or, going further, a researcher could register an analysis and then fabricate the data.

We agree that registration cannot prevent data-fabrication, but do not see this as a critical objection, as preventing data-fabrication is not one of registration's purposes.

4(b) Registering Pre-Analysis Plans on Pre-Existing Data Invites Fraud. With respect to fraudulent claims about the timing of analyses, we expect that outright, intentional fraudulent behavior will be rare (though we have some worries that inviting registration of analyses on historical or pre-existing data might invite a type of fraud). We believe that much data fishing and post hoc rationalization is inadvertent. It is also made easy by the fact that researchers are almost never asked directly to make basic factual claims about the research process -- e.g., about the sequence in which they have carried out research steps. Under current informational conditions, the selective reporting of model results in fact involves making no explicit false claims. We expect that merely asking researchers to publicly make claims about the the timing of data-collection (or analysis) relative to registration would generally elicit honest responses -- i.e., that relatively few researchers would be willing to post outright lies. Moreover, for many forms of research, empirical work involves third parties (e.g., survey firms) or large research teams, allowing the timing of data-collection to be independently confirmed, if necessary.

Note that by creating a registry, with a time-stamp, researchers are forced to go "on record" in a serious way. The public nature of the registry means that there is a permanent record of any claims, available to anyone who might wish to investigate. A lie told once will live forever on the public record and may be contradicted by those with direct evidence (e.g., someone at the specified research site who encountered the researcher prior to the date at which s/he claims to have made his/her first visit). Although the determined researcher can always find ways to fabricate a plausible lie, it would take a good deal of determination and no little risk to do so. We assume that falsifying a registry report constitutes academic malfeasance sufficient to jeopardize tenure and other professional emoluments.

5 Registration provides Scope for Scooping. There is a concern that by registering a research plan, researchers expose themselves to having their ideas stolen and their projects scooped. In

practice, however, all existing registries allow some form of “gating” which allows researchers to keep registered documents private for some reasonable period, such as a year or two. To date we know of no instances in which gated work has been stolen from a registry.

6 Registration is only meaningful for experimental work. Some think that registration is most meaningful for experimental work and may not be appropriate for observational research. The medical journals focus on registration for RCTs, for example, and the AEA registry is limited to RCTs also. The RIDIE and EGAP registries are open to both experimental and observational research. As we explore further below, we believe the more important distinction is between prospective and retrospective research, rather than between experimental and observational research.

7 Registration is only meaningful for prospective analyses. Arguably registration will not be capable of distinguishing exploratory from confirmatory research if what is registered is an analysis of pre-existing data. There are two causes of concern. One is that there is no way to know for sure if an analysis has *already* been conducted prior to the registration. This concern assumes that there are untrustworthy researchers seeking to game the system. We expect that different research consumers will have different views on the trustworthiness of different researchers and that a registry will be informative for historical research as long as it provides clear information on claims about the *timing* of registration. Some readers may discount such registrations entirely, others may find them informative. A second, deeper, problem is that informal analysis always takes place prior to analysis of historical data. For example if past research finds a relation between democracy and growth and a new projects seek to gather better data measuring democracy for the same historical cases, the resulting test with the new data is not an entirely new test since the new data will likely be correlated with data that motivated the project. We believe this is a difficult problem but that it is primarily a general problem about inference from historical data and not a problem specific to registration. Registration may, however, help structure information to make inferences easier, for example by gathering systematic information on what parts of an analysis are new or not -- more on this below.

In any case we note that the function of combatting *publication* bias is supported by registration of such historical work even if it fails in these cases to counter analysis bias. This alone gives some grounds to include this historical work.

8 Registration creates unnecessary bureaucratic hurdles. There are reasons to worry that registration norms will slow researchers down and create unnecessary additional bureaucratic hurdles. These worries resonate with frustrations with current IRB systems. One response to this is that, if properly implemented, registration processes should only seek information that researchers are obliged to generate in course of their research, e.g., a hypothesis and a plan of data collection and analysis. The key difference is that registration requires that some of this be front-loaded. For example, generating analysis code is often done after data gathering but for a fully registered pre-analysis plan this would need to be generated beforehand. Beyond altering

the sequencing of work, completing design elements prior to data gathering has the advantage of alerting researchers to potential weaknesses in their design (such as whether any data needed for analysis has been overlooked) and may shorten the time between implementation and the generation of results.

We recognize however that this can have important implications for time allocation since the decision to invest time in a project cannot be made so easily conditional on results. We believe this may have benefits however in ensuring time investment goes more to projects that are *ex ante* well motivated.

9 Replication is a better use of finite resources to promote credible research. Some have argued that the goals of registration are best met through replication and that this is a better use of limited resources (Laitin 2013). We agree that there should be a lot more replication in political science (whether in the form of reproducing results using an author's data or, more ambitiously, testing a claim with an identical design and new data) and recognize that unfortunately few projects are replicated in political science --- the professional rewards to replication are seen as low --- and that few replications are published or otherwise made public.

Rather than seeing them as substitutes, we see natural complementarities between registration and replication. The greater structure placed on research may facilitate the generation of replication archives. Moreover, a high-quality registration provides a replicator with information on the goals of an analysis and may even provide code and/or decision rules in advance. Finally registration of replications (which are retrospective analyses) may be especially important to reduce the scope for seeking out contrary findings. After all, replication efforts may be especially likely to suffer from publication bias, in which successful replications may be more likely to remain unpublished and unknown. It is also possible that replicated studies suffer from analytic biases: whether similar to those arising in original studies (in favor of positive results) or opposed (in favor of contrary results).

10 Registration will generate inequalities across scholars and methods. Some have voiced concerns that registration will be adopted disproportionately for experimental research, or prospective quantitative research generally, and will be used to privilege this type of research over other modes of research for which registration may not be as suitable.

One response to this concern is to ensure that registration in political science is implemented in a way that it makes it usable across methods. In addition, we note that some of the "caché" accorded to registration derives from a prevailing research norm that privileges testing over exploration and induction. Registration may help cut against this bias by limiting unwarranted claims to have tested a hypothesis and forcing researchers to call valuable exploratory work what it is. It is even possible that quantitative studies in political science are more likely than qualitative studies to make unverifiable claims about "testing" --- indeed the idea of testing is hardwired into the standard statistics published in regression tables. If these claims are called into question - because the researcher has not registered his/her hypothesis and research design, or has

deviated from that design in significant ways, then the result may be to generate a more even playing field (so to speak) between quantitative and qualitative research. Insofar as members of both research communities are doing “soaking and poking,” both should be regarded as exploratory.

INVITATION TO READERS: ADD MORE ARGUMENTS AGAINST IN COMMENTS →

3.3 Impact

Despite the arguments for the positive and negative effects of registries there is to date little evidence that registries have any effect at all. No analyses has been conducted yet of the fledgling social science registries. Studies of the medical registries suggest that registrations very often contain missing or uninformative data and that deviations that yield significant findings are common (Mathieu et al 2009). One study of the distributions of published p values before and after the introduction of registration requirements in 2005 found no evidence of shifts above or below critical thresholds (Humphreys et al 2015). Another study (Kaplan et al 2015) however found evidence that the share of null findings on major studies evaluating drugs or dietary supplements for the treatment or prevention of cardiovascular disease has increased over time.

3.4 Varieties of Research: Experimental vs. Observational, Quantitative vs. Qualitative, Prospective vs. Retrospective

The concept and practice of study registration has been most closely associated with experimental research. Might registration also have uses for observational research? If so, are there certain kinds of observational work in which the gains from registration might be greater or more limited? Could registration be useful for certain kinds of *historical* and even *qualitative* research? We consider here how the possible uses of registration might depend on the type of research being undertaken, proceeding in three steps. First, we briefly address the difference between experimental and observational research as it relates to registration. Then we examine the quantitative-qualitative distinction, focusing in particular on whether registration of qualitative research poses special challenges. Finally, we consider the issues surrounding registration of retrospective as compared to prospective research.

Many of the issues that we explore in this section go beyond the narrow question of how PSDA might function and relate to deeper methodological and epistemological issues surrounding the logic of hypothesis-testing. We intend this discussion to be helpful, however, for thinking through how registration might offer scholarly benefits beyond the realm of experimentation as well as for weighing the merits of establishing a unified political science registry that is open to many forms of empirical research.

3.4.1 Experimental vs. observational research

The key features of experimental work that makes it well suited to registration are both its *prospective* and *manipulative* nature: the fact that scholars can commit to a research design and analysis plan before the data being analyzed have been collected -- and, even better, before the events being studied have occurred. The manipulative nature of the enterprise is critical in that it allows the pre-registered test to be applied to data that are fully independent (conditional on underlying systematic features of the phenomenon being studied) of all data that could have already been observed.

Much observational research, on the other hand, is *retrospective* (i.e., historical), involving the study of events that have already occurred. We discuss below the complexities that arise for registration of retrospective research. Yet it is important to recognize that observational research can also be prospective. Prospective observational designs, for example, might include plans to collect data on an election that has not yet occurred to test a claim about campaign effects, or plans to analyze the impact of an economic crisis that has just occurred on the fortunes of incumbent parties. However because it is not manipulative, data collected on future events that have great continuity may be little different from data collected on past events.

At first blush the non-manipulative nature of observational work may limit the utility of registration. For example an accidental relation between X and Y in time 1 may lead to an observationalist registering a study of X and Y in time 2, knowing that neither quantity is likely to change much and any relation, spurious or substantive, might persist. This seems to stand in sharp contrast to an experimentalist who manipulates X afresh and generates a new test. We believe however that this does not undermine the usefulness of registration; rather, it draws attention to the need for a researcher to clearly identify which part of a design is prospective and which part of an analysis is really novel in order for readers to assess the contribution.

As an aside we note that similarly experimentalists can also engage in a type of fishing despite registration. Say a researcher is interested in the effects of X on Y and examines historical data in 20 sites, and then selects for a new intervention the sites with the greatest historical evidence of effects. The new experiment will teach us something clear about the X, Y relationship in those selected sites, but, without the context of the other sites, we may be misled about how the X, Y relationship might travel across sites. Again here the context and novelty of the test for the general proposition should be clarified in order to evaluate the contribution of the research. Registration alone cannot fulfill that function.

Thus, while experimentalists have shown the greatest interest in registration initiatives, there is in fact nothing intrinsic to experimentation *per se* -- i.e., to the mechanism through which cases are assigned to treatment -- that makes it uniquely suited to registration. In the medical sciences, for

instance, the registration of observational research is gathering pace, with about 18% of all studies registered at clinicaltrials.gov being observational.⁴ The International Initiative for Impact Evaluation's (3ie) registry for development research (RIDIE) also welcomes observational research.⁵ We see no reason why a PSDA could not similarly be open to observational research that seeks to test hypotheses.

3.4.2 Quantitative vs. qualitative research

Registration has to date also almost wholly been associated with quantitative research. Here we focus in particular on the potential uses, limits and challenges of registration for qualitative research.

Qualitative research is very often exploratory. In exploratory research, the scholar seeks to develop a theory in close communication with the evidence. The more back-and-forth there is between evidence and ideas, the more the process of discovery is enhanced. Indeed, the capacity of qualitative research to generate new theoretical insights through close engagement with cases is often considered one of its key comparative advantages. As flexibility, rather than rigidity, is the hallmark of this genre of qualitative research, there may not be much use for registration for purely inductive forms of qualitative research.

Qualitative scholars' interest in induction does not fundamentally distinguish this research tradition from quantitative approaches however, as statistical empirical work also frequently has a strong exploratory component. Moreover, over the last decade, there has been a striking move toward the conceptualization and use of qualitative methods as a set of procedures for testing explanations and theoretical claims. The literature on qualitative process tracing has developed increasingly sophisticated ways of thinking about different types of qualitative tests. Many qualitative methodologists have advocated the use of, and elaborated upon, Van Evera's (1997) typology of tests, which classifies empirical predictions according to the "uniqueness" and "certainty" of the implications being examined (see also Mahoney 2012; Collier 2011; Bennett 2010). This schema yields test types known as "hoop tests," "smoking gun tests," "doubly decisive tests," and "straw in the wind tests," with differing consequences for inferences when passed or failed. For instance, while failure of a hoop test severely damages a theory, the passing of a hoop test only modestly increases confidence in the theory. For a smoking gun test, the consequences are reversed (see also Waldner 2015 for more when counterfactual causal claims are warranted from process tracing). More recently, a number of qualitative methodologists have begun formalizing process tracing tests as forms of Bayesian updating on the probability that a hypothesis is true (Humphreys and Jacobs Forthcoming, Bennett 2015, Beach and Pedersen 2013, Rohlfing 2012).

⁴ <https://clinicaltrials.gov/ct2/resources/trends#TypesOfRegisteredStudies>. See also calls for expanded registration of observational medical research such as Williams et al (2010).

⁵ <http://ridie.3ieimpact.org/>

Qualitative practitioners also commonly claim to employ case study evidence to *test* -- rather than merely to induce -- explanations and theories. To illustrate this point, we conducted a search of all articles appearing from 2010 to July 2015 in 16 highly ranked political science journals that regularly publish empirical work.⁶ Of 92 articles conducting substantive qualitative empirical analyses during this period, 40 -- comprising 43% -- explicitly claimed to have used the qualitative evidence in a confirmatory fashion: i.e., to test or otherwise evaluate the veracity of an explanation, theory, or other claim. While inductive work remains a large focus of the qualitative tradition, it is clear that qualitative scholars on the whole frequently seek to use within-case evidence to test propositions.

This means, however, that consumers and practitioners of qualitative research often face a problem very similar to that confronted by those engaging with quantitative tests. Readers of the 40 "confirmatory" qualitative studies captured by our search will have had little way of knowing whether hypotheses in fact preceded or were induced from the evidence presented. Consider, moreover, the plight of the case-study researcher who *in fact* goes into the field with a clear hypothesis, looking for a specific set of observable implications derived from that hypothesis. Currently, that scholar has no commonly recognized mechanism for credibly communicating to her readers that she did in fact subject her hypothesis to a test.

Moreover, the basic analytical difference between induction and testing is as relevant to qualitative analysis as to quantitative. In quantitative work, the results of a statistical model "fished" from the data can provide important theoretical inspiration, but provide unclear evidence about hypotheses and thus about the theory motivating the model. Likewise, in qualitative work, an explanation induced from causal process observations in a case may suggest crucial new insights into the causal variables or theoretical mechanisms that might be at work. This process is vital to the advancement of research agendas. Yet the clearest *evaluation* of explanatory or

⁶ The journals searched were: *American Journal of Political Science*, *American Political Science Review*, *British Journal of Political Science*, *Comparative Politics*, *European Political Science Review*, *International Organization*, *International Security*, *International Studies Quarterly*, *Journal of Conflict Resolution*, *Journal of Politics*, *Perspectives on Politics*, *Political Research Quarterly*, *Security Studies*, *Studies in American Political Development*, *Studies in Comparative International Development*, *World Politics*. Abstracts were first searched for the terms ("case study" OR "case-study" OR "qualitative" OR "process tracing" OR "process-tracing"). Abstracts of all hits were then examined to determine whether they were making substantive use of qualitative empirical methods (including as part of a multi-method design), yielding 92 qualitative articles for the period. Abstracts of these 92 articles were then read to determine whether the author(s) claimed to have used the qualitative evidence to test (or confirm or find support for or challenge) a theory, explanation, or other claim. Where the abstract was unclear on this point, the article text was inspected. We thank Pascal Doray-Demers for excellent research assistance.

theoretical propositions derives from a new set of observations, independent of those that inspired the propositions in the first place (whether drawn from the same case or other cases).

Thus, in principle, registration of qualitative studies -- and of the plans for analyzing/assessing the evidence collected in light of a set of hypotheses -- could provide a device for scholars interested in implementing process-tracing or other qualitative tests to credibly communicate their research strategy. The intention of a registry open to qualitative work would not be to seek to make testing the dominant, or even a more common, mode of qualitative inquiry; nor would it be to certify testing as a more valued activity. Rather, it would be to provide a new institutional tool for those qualitative scholars who do seek to implement and report tests -- which appears to be quite a lot of them -- to allow scholars and their readers to more clearly differentiate testing from exploration. Indeed, registration could have an especially large signaling benefit in a research genre usually associated with (or assumed to consist entirely of) loose, exploratory work.

We now consider two challenges to using, and reaping the benefits of, registration in a qualitative research context: the problem of precision and the problem of iteration.

The problem of precision in qualitative research

A key element of registration, as we propose it here, is the specification of *pre-analysis plan* (PAP). That is, not only does the researcher indicate the hypotheses being tested and the study design (e.g., experimental treatments, sampling and data-collection procedures, etc.), but how precisely the evidence will be analyzed. In a PAP, a researcher indicates in detail how she will test the propositions or estimate the quantities under consideration. In a quantitative context, a PAP ideally provides the exact measures and model specifications to be employed, possibly even providing the code to be used for the analyses and mockups of the tables in which results will be presented.

In a qualitative context, researchers could likewise specify in advance a great deal about how they plan to implement and evaluate their tests. Qualitative scholars could state in advance the theory(ies) or explanation(s) being examined; identify the cases being studied; spell out the observable implications of those theories that will be tested in these cases; specify the kinds of evidence or observations that, if found, would satisfy those predictions; and state how passing or failing such tests would affect the validity of the theory. We might consider the last three elements the key components of a qualitative PAP.

At the same time, in the absence of metrics such as beta coefficients and p-values, it will often be hard to specify in advance exactly what form of qualitative evidence counts as satisfying an empirical prediction. For instance, a qualitative scholar might seek to explain tax cuts adopted by a conservative government, and might want to test the proposition that the tax cuts were motivated by an interest in Keynesian demand management. The scholar might further specify a "hoop test" along these lines: when I go into the archives and examine the record of deliberations among cabinet officials about the tax cut, I should see prominent mention of the logic of

Keynesian stimulus. To say this a “hoop test” is also to specify in advance that failure to find such evidence would be highly damaging to the explanation. This is already committing to quite a lot, but it of course still leaves some things open. Just how prominent do mentions of Keynesian logic have to be for the test to be passed? How many actors have to mention it? What forms of words will count as the use of Keynesian logic? Qualitative data-search procedures may also be harder to specify in advance than quantitative sampling procedures, especially given the common difficulty in identifying the “sampling frame” *ex ante* for common forms of qualitative evidence-collection, such as archival research and elite interviewing.

The ambiguities of theory-testing in a qualitative mode must be reckoned with by any researcher who undertakes this style of analysis. Importantly, however, *the use of a registry does not add to these difficulties in any respect*. The reviewer or reader of a registered qualitative study would ultimately have to make a judgment about the extent to which the evidence presented constitutes the satisfaction of the empirical prediction as specified in the study’s PAP, just as readers of unregistered qualitative studies must judge the persuasiveness of the empirical account. But registration offers the possibility that such judgments could at least be made with clarity about which came first: test-specification or test result.

Moreover, in a world in which registration is an option, the equilibrium outcome is likely to be one that substantially mitigates the problem of precision in qualitative testing. Audiences will be free to judge how much “wiggle room” the researcher has given herself in her PAP, and to assess the credibility of the test result accordingly. In turn, then, qualitative scholars interested in hypothesis-testing will be incentivized to bind their own hands as effectively as possible -- i.e., to come up with maximally clear pre-specifications. A further possibility is that qualitative researchers might turn more to devising tests for *one another’s* hypotheses, which should further sharpen incentives to specify “brittle” predictions.

In sum, we see the problem of precision in qualitative testing as an operational challenge that scholarly communities could work out for themselves, rather than as a fundamental challenge to the usefulness of registration for qualitative work.

The problem of iteration

A further challenge arises from the complex mixes of induction and testing characterizing much qualitative research. Qualitative scholars often iterate between exploration and the evaluation of hypotheses. A case-study researcher might, for instance, (a.) articulate some initial hunches derived from the literature; (b.) engage in some “soaking and poking” in the field, which inspires the elaboration of more refined hypotheses; (c.) derive observable implications of those hypotheses, (d.) look for causal-process observations to test these implications; (e.) find the hypotheses only partly supported and amend them to better fit the evidence; (f.) derive new observable implications of the amended hypothesis; and then (g.) look for additional data to evaluate these amended predictions.

Such iteration raises no fundamental problem for registration. Registration need not take place prior to all empirical work on a project. Rather, tests could be registered at any point in the project's development, as long as there is some discrete set of evidence that has not yet been observed and that the analyst plans to seek and engage with specific hypotheses and observable implications in mind. Thus, for instance, in the above iterative sequence, a scholar might register a set of tests between stages (c.) and (d.), and/or between stages (e.) and (f.): after a period of inductive theory-development and test-derivation and just before the collection of data to implement the test. Thus, for instance, the tests carried out at stage (d.) could be conceived of and registered as a discrete "study" (while remaining part, of course, of a larger project) with its own design and pre-analysis plan.

At least two elements of the logic of registration and credible testing become more complicated, however, with iteration. The first is establishing exactly when the data collected at stages (d.) and (f.) have in fact been collected, particularly if a scholar has spent the entire period or several spells in the field, observing the phenomenon of interest. Were the test data in fact observed only after the test plan was registered? The second complication arises in establishing the independence of the evidence used in the test and the evidence used to induce theoretical insights. We explore both of these problems further in the next section.

3.4.3 Prospective vs. retrospective (historical) research

More fundamental difficulties arise for registration of studies that are retrospective, rather than prospective: that is, those analyzing data on events that have already occurred. We can think of the problem as taking two forms: a problem of prior observability and a problem of informational independence.

Prior observability

Many of the key benefits of registration hinge on the researcher's not having seen the test data prior to registering the design and its PAP. We refer to this situation as one of *prior unobservability* of the test data.

In experimental work, establishing prior unobservability is relatively straightforward because outcome data are gathered only after treatments are administered (assuming that registration occurs prior to that administration). Experimental research is in this sense always prospective, involving the observation of events that have not yet occurred. As long as the date of experimentation can be established, one can know that hypotheses and tests registered prior to that date provide confirmatory evidence.⁷

⁷ We are discounting the possibility that the experimentalist routinely runs unregistered experiments in the lab, then, after the fact, only registers the design yielding desirable outcomes.

Likewise, in the realm of observational research, prior unobservability can be readily established wherever the events being studied have not yet occurred. Consider a researcher who will be testing a hypothesis about campaign dynamics using the next round of data from the American National Election Study. Prior unobservability can be achieved here by registering a hypothesis and research design prior to the election campaign.

At the other end of the spectrum, much retrospective research involves data that already exist. The election study has already been carried out, written about, and its dataset archived; the newspaper stories to be analyzed have already been published; the cross-national indicators to be used have already been compiled. Where the data to be analyzed come from a dataset that already existed in analyzable form prior to registration, establishing the prior of unobservability of the evidence may effectively be impossible. While study registration in such situations may have other uses (e.g., allowing opportunities for comment on study designs, providing a clearinghouse for the reporting unpublished results), registration will be a weak tool for credibly establishing that testing has taken place.

However, a credible case for prior unobservability can be made for retrospective research wherever the data themselves are newly collected, newly published, or otherwise newly available, and hence (prior to that point) inaccessible to the researcher. Opportunities of this sort arise prior to the implementation or release of a new survey, the opening of a new archive or a new collection of historical records, the discovery of new sources (e.g., from an archeological site), and so forth. For instance, the researcher who plans to conduct a survey measuring civic activity may effectively be measuring activities that have already occurred; but it may be straightforward to document that the survey itself was not fielded until after the date of registration. In principle, any documents obtained through a Freedom of Information Act (or similar) request for undisclosed documents provide an opportunity for establishing prior unobservability. Further, any dataset whose purpose and structure is known to the research community prior to the publication of measures also provides an opportunity for registration with prior unobservability. For example, the Varieties of Democracy (V-Dem) project will release its first round of data (for full downloads) on 31 December 2015. Prior to that point, researchers might register a research hypothesis and analysis plan, building on the full V-Dem codebook (available at V-Dem.net).

Thus, wherever data-collection is a demonstrably prospective activity, it may be possible to establish the prior unobservability of the evidence. Importantly, much qualitative work is well suited to establishing prior unobservability insofar as qualitative research very often involves original data collection, whether in the form of interviews, focus groups, participant-observation, archival work, or the original coding of some feature of the world. Moreover, the collection of qualitative data frequently involves the researcher undertaking a discrete activity -- e.g., a visit to the archives, an elite interview, a field-site visit -- with a specific start date. For many qualitative projects, there is thus a relatively clear point in time prior to which the evidence being analyzed was *ex ante* unobservable to the researcher. Perhaps more so than many researchers using

quantitative secondary data, qualitative scholars may have quite rich opportunities to register study designs, hypotheses, and analysis plans prior to encountering their data.

Credibility problems might arise. It is always a possibility that a researcher might lie about when data were collected or came into her possession. This problem as it arises for retrospective research, however, is not fundamentally different from the problem of establishing when experimental treatments were administered or outcomes measured. Moreover, we can imagine a range of devices through which scholars' claims about the timing of qualitative data-collection tasks might be made independently verifiable. Archival visitor logs, third-party research-service providers (e.g., survey firms), and partner organizations, for instance, could provide a source of verification. Doctoral supervisors could certify the timing of their students' fieldwork visits; for established scholars, colleagues might be called upon to do so.

None of these mechanisms would be airtight. While the time of a field visit can be established, for instance it is harder to establish the *absence* of a prior field visit. Importantly, however, it would not be the registry's task to delineate the appropriate mechanisms for establishing the timing of fieldwork. Rather, the details of these devices would be deliberated upon by the relevant research communities, would be at the discretion of individual researchers, and would of course be open to judgment by reviewers and readers. We expect, moreover, that once the principle of registration becomes recognized by the academic community, researchers undertaking theory-testing work will have incentives to develop novel and increasingly credible procedures for establishing the prior unobservability of their test data.

We further note that the challenge of establishing prior unobservability is not a problem of registration per se, but of any claim to empirical testing. Registration eases the problem by, minimally, establishing a date on which hypotheses, study designs, and/or analysis plans were in place.

Informational independence

A deeper problem of retrospective testing relates to the possible dependence of "new" data on old. Even if researchers have not yet observed the data they are analyzing, they have usually observed other things about the historical events or cases being studied; and what they have observed may be *correlated* with what the test data that they plan to collect. For instance, in the case of survey research, a scholar may not have fielded her questionnaire at the point of registration, but she may have spent time in the social context being studied, read news reports about public sentiment, read other social analyses of the case, etc. In other words, she may have already made observations that are correlated with the survey measures she will be collecting and using to test her hypotheses. Her hypotheses and test specifications are thus likely to be shaped by information that is related to her test measures. Similarly, even if a scholar has not yet been to the archives to read particular documents, she has possibly read histories that reveal facts that are correlated with the new facts she will observe in the archives.

Data about past events may never be perfectly independent of observations already made. There will nevertheless be variation in how independent the test data are of prior information, and greater independence should yield clearer tests. We note that the issue of informational independence could also potentially affect prospective, and even experimental, work, with independence declining the later in the study implementation process registration occurs.

3.5 Informational Stages of Registration

We believe that many of these features such as prior unobservability and independence are critical for evaluating research and also important for determining in what ways a registry is useful for different types of research.

We see the chief merits of a registry as being in the provision of transparency and believe that the best way that a registration process can handle these different features of a design is in ensuring that they are articulated as clearly as possible.

Table 2: Informational Stages of Study Registration

| Prospective Research | Retrospective Research |
|---|---|
| P1: Before any research starts | R1: Before the collection of any data relevant to phenomenon of interest on any cases in study population |
| P2: Before units are assigned to treatment (whether by the researcher or by nature) | R2: Before the collection of data on the specific measures to be used in the planned analysis for any cases in study population |
| P3: Before treatment has been administered (and thus before outcomes could have occurred) | R3: Before collection of any data on the specific cases being studied |
| P4: Before data collection on outcomes | R4: Before collection of data on the specific measures to be used in the planned analysis for the cases being studied |
| P5: Before researcher access to any data | R5: Before researcher access to specific data to be used for analysis |
| P6: Before any analysis of outcome data | |

Clarifying the meaning of tests via registration is simplest for registration prior to the start of a fully prospective study. Yet registration can also help audiences interpret tests of all kinds in more complex research situations. Researchers registering their studies might maximize these

signaling benefits by indicating the stage in a project or collective research program at which registration is taking place. Table 2 illustrates a conceptualization of stages in the research process at which different kinds of prior information have already been observed.

As can be seen in the table, the degree to which observations are independent of past observations tends to decline as one moves down the table (though the progression is likely not perfectly monotonic). For prospective studies, for instance, a researcher who knows which units have been assigned to which treatment status (as at Stage P3) may have more information predictive of possible findings than a researcher who doesn't: this may be especially true if he has other information about the units that may be predictive of their outcomes. Registration at Stages P4 through P6 -- which all take place after treatments have been administered and thus outcomes could already have occurred could be seen as the registration of retrospective analysis if one did not trust the analyst and this lack of trust would generate the same issues of prior observability and informational dependence as we would see in other forms of historical research. We should note that we believe that the default position of the research community is to trust others. Thus, for example, if a scholar claims not to have access to the outcomes --- even if the outcomes have been collected --- we advocate treating such a registration as prospective.

For retrospective research, the researcher at R1 -- where nothing whatsoever has yet been observed about the historical events of interest in any case -- is essentially in the same position as a prospective researcher. Moving down the column, seeing data on cases outside the study sample may reduce test strength to the extent that the out-of-sample cases may or may not be independent of the in-sample cases. The greatest informational dependencies likely arise, however, as researchers are exposed to data from the same cases that they plan to analyze, which may be correlated with the data they plan to collect and test hypotheses against.

For each stage, questions of informational independence and prior unobservability arise. For instance, at Stage R4, the researcher will likely want to make an argument -- whether theoretical or empirical -- about how independent the yet-to-be-collected test data are of observations previously made in the same case. Audiences' judgments of that claim will condition their assessments of the strength of the test. At the same time, the researcher will want to establish that *in fact* the specific measures to be analyzed had not yet been observed (i.e., that registration did in fact take place at Stage R4).

Clearly, the interpretability of reported tests will vary with the stage at which the research is registered. Some readers may be inclined to view tests as credible only when registered at the very early stages -- or perhaps only for P1 registrations -- while other audiences may be willing to be persuaded of the prior unobservability and partial independence of the test data for historical research. Researchers themselves may also devise increasingly sophisticated ways of establishing these test criteria. Across all of these scenarios, registration itself will make both researchers' tasks and readers' judgments easier by establishing the point in time at which study designs and tests were specified.

4 Existing options for registration for Political Science

A number of registries or proto-registries exist already and some of these already fulfill some of these functions. These include registries established by the *American Economic Association* (for RCTs), by the development research funder 3ie (for research on program evaluation in international development) and by the *EGAP* Network. In addition, other multi-purpose platforms such as *Dataverse* or the *Open Science Framework* can be used for registration purposes.

Table 2 provides information about these registries, their activity and the facilities they provide.

Table 3: Current Registration Options

| | AEA LINK | DATAVERSE LINK | EGAP LINK | OSF LINK | RIDIE LINK |
|--|-----------------------------|-----------------------------------|------------------------------|-----------------------------|-------------------------------|
| DOMAIN: | RCTs | OPEN | Governance and politics | OPEN | Development interventions |
| # Studies Pre-registered / Registered [20150725] | 82 / 421 | 2 / 2 | 104*/ 183 | ? / 1311 | 64 / 64 |
| Minimal: | | | | | |
| 1 A lock box | ü | ü | ü | ü | ü |
| Core: | | | | | |
| 2 Provide Identifiers | ✓ | ✓ | ✓ | ✓ | ✓ |
| 3 Versioning | ✓ | ✓ | ✓ | ✓ | ✓ |
| 4 Minimal design confirmation | ✓ | | ✓ | | ✓ |
| 5 Searchability (Between Registries) | (✓) | | | (✓) | |
| Additional: | | | | | |
| 6 Control public access. | ✓ | | ✓ | ✓ | ✓ |
| 7 Access anonymized versions | | | (✓) | ✓ | |
| 8 Connect designs to studies and data | | (✓) | (✓) | (✓) | |
| 9 Allow commenting | | | | ✓ | |
| 10 Registration confirmation | | | | | |

Note traffic information as of 25 July 2015. There are plans in place to make the AEA and OSF registries inter-searchable though these plans are not yet in place. Dataverse and OSF allow registrants to cross link to data within their platforms in an unsupervised way. The EGAP registry seeks to link update entries with links to published research and materials as this becomes available. OSF has the largest number of registrations, however since this platform is unsupervised there is no guarantee that these registrations in fact include registered research designs.

* Estimate as precise data on timing of registration relative to project progress is available for 145 studies only. Of these, registration took place prior to any research activities in 33% of cases, prior to treatment assignment in 57% of cases, prior to realization of outcomes in 70% of cases, and prior to researcher access of data in 84% of cases.

Currently none of these are “full service” registries. Dataverse and OSF are unsupervised registries which means that they have the advantage of running with little infrastructure yet still provide basic functionality. They do not however put constraints or structure on the contents of the registrations. To date the Dataverse platform has attracted few registrations. The AEA and RIDIE registries are open to a limited set of study types but provide basic supervision and good internal search functionality. The EGAP registry has had reasonable uptake among political scientists, it provides basic supervision, some searchability and some linking of registrations to supplementary information. It does not allow commenting and does not provide confirmation of registrations. The EGAP registry has perhaps the greatest focus currently for political science research, though the association with EGAP which focuses on experimental research largely in developing countries, may ultimately limit the scope of this registry.

Of those supervised registries that collect minimal data (AEA, RIDIE, EGAP), none have been tailored to gather data of particular relevance to qualitative or historical research.

5 A Proposal

The three methodology sections of the American Political Science Association should seek funding to establish a *full service* registry for political science guided by a set of six core principles.

5.1 Principles

The guiding principles of the political science research registry should be

- **Transparency:** A primary goal of the registry is to enhance transparency regarding what research is undertaken and how it is undertaken. The purpose is not to constrain researchers or to privilege one type of research over another.
- **Neutrality:** The registry should not be owned by a particular part of the discipline or used to support a particular method. It should be open to multiple types of research in terms of method and substance. In particular it should be welcoming of experimental, observational, qualitative, and historical research.
- **Navigability.** The registry should make it easier to find research undertaken in the discipline and to connect research designs with research findings, data, and other supplementary materials.
- **Complementarity:** The registry should complement other transparency initiatives in political science. In particular it should facilitate journals seeking to provide anonymous registrations to reviewers, and to confirm that studies or analyses that authors claim are registered are indeed registered. It should facilitate efforts by journals and others to

publish “registered reports” -- which allow for commenting on designs prior to implementation.

- **Automated:** To the extent possible the functions of the registry should be automated and should not rely on human judgement. A particular challenge is to develop a method so that the CAP function can be implemented in an automated way. More on this below.
- **Permanence:** The registry should be established such that registered data be permanently archived and accessible even if the registry itself is short lived or becomes superseded by an alternative institution.

5.2 A Full Service Registry

The registry should provide all ten functions as described in Table 1 above. For the case of political science, and given the considerations described above, we believe four of these functions require special consideration. In particular:

Minimal Supervision: We see three components of minimal supervision beyond current practice.

- First in addition to fields commonly present in existing research registries, registration forms should include fields specifically crafted for observational and historical analyses. In particular this should include more granularity of information regarding the timing of research consistent with the categories in Table 2 above.
- Second, to allow the sharing of anonymous registrations with journals, supervision should also verify the anonymity of files declared to be anonymous.
- Third, for PIs seeking to register CAP *pre-analysis plans that they seek to have validated later*, and for an enumerated set of computer languages, the registry should verify that that pre-analysis code indeed produces tables as indicated. To ensure a minimal workload registrants could be asked to provide analysis plans in the form of a markdown file or equivalent. (See [here](#) for an example of a preanalysis plan that includes functioning code).

Searchability: Two innovations may enhance searchability. First designs registered with the political science registry could be cross posted to the Open Science Framework (OSF) which would ensure searches include searches of OSF designs as well as AEA projects. Second since political science does not have an analogue to the JEL codes, the registry could develop over time a set of keywords to facilitate research classification, starting from current classifications such as the [BISAC](#) codes.

Commenting: Allowing the possibility to comment on registered designs is one way to enable that weaknesses in an analysis plan are found prior to implementation of research. Commenting however also runs the risk of producing uninformative or even destructive dialogues. We propose

that PIs have the option to allow commenting, that comments should be minimally moderated to ensure civility, and that PIs should retain rights to remove comments posted on their designs, have the option to turn off comments, and assume the responsibility of primary moderator.

Confirmation. Perhaps the most difficult innovation is the confirmation of registration, COD and CAP. The purpose of confirmation is to find a way to signal to consumers of research that there has been a meaningful registration without imposing additional burdens on readers or journal editors. Many complexities arise in establishing whether a piece of research is really registered or not however. First, it is possible that a design is registered but that an ultimate analysis bears little resemblance to the declared analysis. It is also possible that multiple analyses are registered but that only a subset of these enter an analysis; or conversely that a study consists of a mixture of registered and unregistered analyses. Adjudicating in these instances runs the risk of creating significant labor requirements for a registry. At first glance, such adjudication might seem to give grounds for concern (described above) that a registry becomes an unaccountable gatekeeper rather than a simple facility.

One approach to confirmation leaves the question in the hands of interested reviewers and readers. Under this approach authors provide a link to registrations. These registrations may or may not be checked by journals, readers, or reviewers. For example, authors might include in a first footnote with simple text of the form:

“The analyses in section 4 were registered after data collection but prior to analyses at www.psda.org with ID (DOI): 4317531. Appendix B describes deviations from the registered analysis plan.”

Note as in this example, journal policy could prescribe details of the footnote, such as requiring that text indicate if registration covers only parts of articles or in which there are deviations from registered designs, this may be also be signalled in short form in the first footnote:

An alternative approach, which we advocate, is to rely on confirmation by the registry. For registry confirmation, the registry could issue COD and CAP certification.⁸

COD: Study registry certification would be provided by the registry to confirm that minimal study registration was completed and corresponds in essential details to the study reported; listing of authors general topic, study site, and general hypotheses or motivation. COD certification is deliberately minimal and used only to prevent obvious abuse of the registry --- such as researchers claiming to have registered a study when they have not.

CAP: Confirmation of Pre-analysis plan certification would be provided if researchers provide a complete description of analysis procedures and a description of any deviations from planned

⁸ Note that the confirmation requirements are less stringent than those proposed by the Open Science Framework for pre-registration badges (see here).

analysis. (Note that for simplicity, confirmation would not extend to other features of a design, such as implementation or measurement.) One simple approach to confirmation of quantitative analysis that removes value decisions is to request that authors provide a registry with a “**reconciliation file**” that shows the differences between a planned analysis and a final analysis, with justifications. It is the reconciliation file that is then confirmed and becomes part of supplementary material. This requires full creation of code in advance, and would make CAP, we expect, a rare event.

The confirmation then simply requires (a) a check that the description of the planned analysis in the reconciliation file matches the registered document (b) a check that the final analysis file implements the final analyses and (c) there are comments describing deviations. Note task (b) may be done by some institution other than the registry, as selected by journals. In the example below the Reconciliation table is provided by authors. From this it is easy to extract the reported planned analysis and compare this with the registered analysis, and the reported final analysis and compare this to the final analysis.

Task (c), in addition, is critical to minimizing the role for unaccountable judgments by validators. The key decision would *not* be some determination of whether a study adhered sufficiently closely to its pre-analysis plan. Rather, confirmation would involve a check of whether any departures from the registered plan are clearly *identified* in the final text.

To facilitate confirmation we propose that the ***Unit of Registration*** should be the *publication* (a research article or book), not the research project. Any analyses in an article that are additional to the registered designs should be included in the comparison document.

Procedures for confirmation of qualitative registrations: While in principle the above procedures could apply to registered studies taking any methodological approach, we do not see registry certification as a viable approach for most qualitative studies. We do not believe that the specification of most qualitative research designs or process-tracing tests can be specific enough to allow for the kind of easy mechanical reconciliation described above. For any analyses that do not take a coded algorithmic form, we therefore recommend a minimalist approach in which any confirmation judgments are left to reviewers, editors, or readers. Manuscript submissions for registered qualitative studies would, minimally, contain links to anonymized study registrations, allowing reviewers and editors to compare study design and pre-analysis plans to the data-collection and tests reported in the submission. Editors could require that deviations from registered designs or analysis plans judged to be significant by reviewers are clearly commented in the article. Undertaking this confirmation places a greater burden on reviewers and editors, which may be worth consideration. An alternative, lower-cost approach would be to simply link the published article to the study registration and leave judgments of tests in the hands of readers. These are, of course, operational details that each journal could work out on its own terms.

5.3 Set-up and Staffing

A registry requires both initial set up work to create a platform capable of implementing the functions outlined in Table 1 as well as staffing.

We believe that most of the technical needs can be built on the platforms already established by the Open Science Framework (OSF). Using the OSF the registry can generate its own front end to an OSF databased with required fields and then approve submissions subject to minimal review. Designs registered in this way would produce an OSF project page that allows for commenting, updating, and versioning. Integrated into the overall OSF structure this allows for searchability with other OSF registrations as well as AEA registrations.

Integration with the OSF platform takes a political science registry very far; yet staffing will still be important. The core function of registry staff would be to (a) conduct minimal supervision and approve registration requests (b) maintain project pages with updated project data (c) undertake minimal moderation of public commentary on research designs (d) provide registration certification to journals as requested. In the medium term we would like to see these functions mostly automated. We estimate a two person staff may be sufficient to oversee a functioning registry for political science. Staff would be employees of the hosting organization, which might be one of several bodies including APSA or IQMR, or perhaps if the registry expands on current initiatives, Dataverse or EGAP.

5.4 Institutional Considerations

We propose that the registration be jointly owned by the sections. Oversight of the registry should be assured by a board consisting of (a) representatives of the three sections, (b) a representatives of the APSA council, (c) three representatives of other APSA sections (d) a representative from an external research body such as the SSRC.

We propose that the three sections seek funding from external groups interested in enhancing research transparency and integrity in order to provide 3-4 year start up funding. Afterwards the registry could explore whether there are ways to introduce a registration fee structure that does not exacerbate existing inequalities⁹ or, if the registry has proven utility for the discipline, this may justify subsidization by the Association at a future point in time.

⁹ User costs are unlikely to be able to cover full costs of a registry. For illustration if costs run at \$150,000 per year for 500 new registrations at steady state, a \$100 user fee would only cover a third of needed costs.

References

- Beach, Derek, and Rasmus Brun Pedersen. 2013. Process-Tracing Methods: Foundations and Guidelines. University of Michigan Press.
- Bennett, Andrew. 2010. "Process tracing and causal inference." In Rethinking social inquiry: Diverse tools, shared standards, ed. David Collier and Henry E Brady. Lanham, MD: Rowman & Littlefield.
- Bennett, Andrew. 2015. "Appendix: Disciplining Our Conjectures: Systematizing Process Tracing with Bayesian Analysis." In Process Tracing in the Social Sciences: From Metaphor to Analytic Tool, edited by Andrew Bennett and Jeffrey Checkel. New York: Cambridge University Press, 276-298.
- Collier, David. "Understanding process tracing." *PS: Political Science & Politics* 44.04 (2011): 823-830.
- Franco, Annie, Neil Malhotra, and Gabor Simonovits. "Publication bias in the social sciences: Unlocking the file drawer." *Science* 345.6203 (2014): 1502-1505.
- Gerber, Alan S., et al. "Publication bias in two political behavior literatures." *American Politics Research* (2010).
- Gerber, Alan, and Neil Malhotra. "Can political science literatures be believed? A study of publication bias in the APSR and the AJPS." annual meeting of the Midwest Political Science Association, April. 2006.
- Humphreys, Macartan, and Alan Jacobs. Forthcoming. "Mixing Methods: A Bayesian Approach." *American Political Science Review*.
- Humphreys, Macartan, Raul Sanchez de la Sierra, and Peter Van der Windt. "Fishing, commitment, and communication: A proposal for comprehensive nonbinding research registration." *Political Analysis* 21.1 (2013): 1-20.
- Kaplan, Robert M., and Veronica L. Irvin. "Likelihood of Null Effects of Large NHLBI Clinical Trials Has Increased over Time." *PloS one* 10.8 (2015): e0132382.Laitin, David D. "Fisheries management." *Political Analysis* 21.1 (2013): 42-47.
- Mahoney, James. "The logic of process tracing tests in the social sciences." *Sociological Methods & Research* 41.4 (2012): 570-597.
- Mathieu, S., I. Boutron, D. Moher, D. G. Altman, and P. Ravaud (2009, September). Comparison

of Registered and Published Primary Outcomes in Randomized Controlled Trials. *JAMA : the Journal of the American Medical Association* 302 (9), 977{84.

Rohlfing, Ingo. 2012. *Case Studies and Causal Inference: An Integrative Framework*. New York: Palgrave Macmillan.

Van Evera, Stephen. *Guide to methods for students of political science*. Ithaca: Cornell University Press, 1997.

Williams, Rebecca J., Tony Tse, William R. Harlan, and Deborah A. Zarin. "Registration of observational studies: is it time?." *Canadian Medical Association Journal* 182, no. 15 (2010): 1638-1642.

Appendix A: Illustration of analysis confirmation strategy

One simple approach to confirmation of quantitative analysis that removes value decisions is to request that authors provide a registry with a “reconciliation file” that shows the differences between a planned analysis and a final analysis, with justifications. It is the reconciliation file that is then validated and becomes part of supplementary material. The verification then requires (a) a check that the description of the planned analysis in the reconciliation file matches the registered document (b) a check that the final analysis file implements the final analyses and (c) there are comments describing deviations. Note task (b) may be done by some institution other than the registry, as selected by journals. In the example below the Reconciliation table is provided by authors. From this it is easy to extract the reported planned analysis and compare this with the registered analysis, and the reported final analysis and compare this to the final analysis.

Note that in practice this places the burden on researchers to demonstrate consistency in an easily verified way (though much of the work may be done in an automated way using tools such a [Compare It!](#)). It also requires the registry staff have the ability to implement the code used and so in the short run any confirmation service might be restricted to say R and stata.

The main judgement call required here is to determine whether there is a genuine rationale provided for all deviations (whether or not the rationale is compelling). For example a reconciliation document could always be created by simply appending an entirely commented out analysis plan to a new analysis, together with a single uninformative justification. We leave this judgement call to research consumers.

Sample Reconciliation Document

| Comp | Plan | Final |
|------|------|---|
| 1 | 7 | 4 # Analysis |
| 2 | 8 | 5 M1 <- lm(Y1~X, data = data) |
| 3 | | |
| 4 | | # Modification 1: M2 changed to include possible confound as control |
| 5 | 9 | # M2 <- lm(Y2~X, data = data) |
| 6 | | 6 M2 <- lm(Y2~X + Z, data = data) |

```

7           # Modification 2: model M3 added to account for multiple
8           comparisons

9           7   M3 <- lm(((Y1+Y2)/2)~X , data = data)

9           10      8

10          11      9   # Results

11          12      10  Out <- cbind(coef(summary(M1))[2, 1:2],
12                           # Modification necessitated by Modification 2

13          13      11   #           coef(summary(M2))[2, 1:2])

14          14      12   coef(summary(M2))[2, 1:2],

15          15      12   coef(summary(M3))[2, 1:2])

16

17           # Summary of modifications

18           # Modification 1: M2 changed to include possible confound
18           as control

19           # Modification 2: model M3 added to account for multiple
19           comparisons

```

This reconciliation document can be verified by checking its two components against the registered design and the final analysis code.

1 Analysis Plan

```

1   # Dummy data

2   X <- runif(10)

3   Y1 <- X + rnorm(10)

4   Y2 <- X + rnorm(10)

5   data <- data.frame(Y1, Y2, X)

6

7   # Analysis

8   M1 <- lm(Y1~X, data = data)

```

```
9 M2 <- lm(Y2~X, data = data)
10
11 # Results
12 Out <- cbind(coef(summary(M1)) [2, 1:2],
13                 coef(summary(M2)) [2, 1:2])
```

2 Final Analysis

```
1 # Import data
2 data <- read.csv("mydata.csv")
3
4 # Analysis
5 M1 <- lm(Y1~X, data = data)
6 M2 <- lm(Y2~X + Z, data = data)
7 M3 <- lm(((Y1+Y2)/2)~X , data = data)
8
9 # Results
10 Out <- cbind(coef(summary(M1)) [2, 1:2],
11                  coef(summary(M2)) [2, 1:2],
12                  coef(summary(M3)) [2, 1:2])
```