Can the Biomedical Research Cycle be a Model for Political Science?

Can the Bio-Medical Research Cycle be a Model for Political Science?

Forthcoming in *Perspectives on Politics*, December 2016

Evan S. Lieberman (Department of Political Science, MIT)
evanlieb@mit.edu

ABSTRACT:

In sciences such as bio-medicine, researchers and journal editors are well aware that progress in answering difficult questions generally requires movement through a research cycle: Research on a topic or problem progresses from pure description, through correlational analyses and natural experiments, to phased randomized controlled trials (RCTs). In bio-medical research all of these research activities are valued and find publication outlets in major journals. In political science, however, a growing emphasis on valid causal inference has led to the suppression of work early in the research cycle. The result of a potentially myopic emphasis on just one aspect of the cycle reduces incentives for discovery of new types of political phenomena, and more careful, efficient, transparent, and ethical research practices. Political science should recognize the significance of the research cycle and develop distinct criteria to evaluate work at each of its stages.

Acknowledgments:

Thanks to David Collier, John Gerring, Kosuke Imai, Jeffrey Isaac, Robert Keohane, Markus Kreuzer, Julia Lynch, Philip Martin, Nina McMurray, Ben Morse, and Yang-Yang Zhou for helpful comments and suggestions.

Introduction

A research cycle in a scientific discipline is constituted by researchers working at various stages of inquiry, from more tentative and exploratory investigations to the testing of more definitive and well-supported claims. As a particular research area matures, scientists are less frequently surprised by new phenomena because core processes are well understood. They are more likely to focus their efforts on making precise estimates of causal effects, often through randomized experiments. And indeed, such a pattern is evident in bio-medical research. In fact, descriptive and correlational work is often published in the major bio-medical research journals, although different criteria are used to assess their significance than are used to assess experimental research.

In this essay, I consider the value of this model for political science. My motivation is a sense that political scientists may be paying disproportionate attention to studies that focus on the precise estimation of causal effects to the exclusion of other types of complementary research – so much so that the range of questions to which we can eventually apply sophisticated strategies for causal inference may become severely limited, curtailing our collective contributions to useful knowledge. Put another way, might we be able to answer a greater number of causal questions, with respect to a richer set of subject areas, if we created more intellectual space (particularly in leading, peer-reviewed journals) for high quality scholarship that does not make strictly causal claims? Or that draws more tentative conclusions about causal relationships? Would scholars be more likely to accurately describe the nature of their contributions if they were not under pressure to report their findings as "causally well identified?"

Specifically, I highlight the need for the type of research cycles and division of labor[1] one sees in other scientific fields, including the bio-medical sciences. The notion of a research cycle that I have in mind is one that is constituted as a scholarly conversation through peer-reviewed publications, and includes a mix of inductive and deductive theorizing and observation. It explicitly recognizes differences in the state of research within a substantively-delimited cycle, such that we might expect a move from more tentative to more definitive claims about causal relationships. It is a *cycle* because we rarely expect research to "end," but merely to generate new observations and surprises that will spur new inquiries. Within a discipline that takes the notion of research cycles seriously, the criteria for what would constitute a contribution to knowledge depends on *where* in the research cycle an author was attempting to publish. More exploratory research could be recognized as "cutting edge" if it were breaking open new questions through identification of novel patterns or processes. Large-scale randomized controlled trials (RCTs) would be more appropriate as a cycle matures, and could provide definitive evidence about more narrowly-defined questions about specific

[1] In a complementary manner, Gehlbach (2015) argues for a methodological division of labor in political science.

causal relationships.

My point is *not* to eschew interest in causal questions or causal relationships, or to challenge the potential value of experimental research in political science. On the contrary, I raise the analogy to bio-medical science as a science that – given its immediate practical search for knowledge to improve human wellbeing, and recognition of the harms associated with faulty inference – is seriously concerned with establishing clear cause-and-effect relationships. But as I will illustrate, causal analysis is just one part of the division of research-oriented labor.

In the remainder of this essay, I begin by describing the manifestation of the research cycle in the publication of bio-medical research, and highlight the extent to which an analogous cycle seems far more limited in political science, at least within leading publication outlets.[2] Subsequently, I propose a framework for developing such a research cycle in the discipline, detailing some standards of

[2] To be sure, the very notion of a research cycle is not new, including within the discipline of political science. For example, Munck (1998) usefully re-framed the central lessons of King et.al.'s (1994) *Designing Social Inquiry* in terms of a research cycle that moves from more inductive observation to hypothesis testing. My point is that there is little evidence that important steps in such cycles are thoughtfully considered, especially through publication.

excellence for evaluation and publication.

The bio-medical research cycle

Bio-medical research and political science differ along many key dimensions such that one might question the utility of using the former as a model for the latter. A great deal of bio-medical research is rooted in a clear and focused mandate to try to develop best practices and new technologies for improving the health of humans, and much research drives product and protocol development. By contrast, a much smaller share of political science research is intended for "practical" ends, as most scholars search simply for deeper understanding of the socio-political world in which we live, and generally have more modest expectations about how research might be used for practical purposes.

In terms of occupation, many bio-medical researchers have responsibilities as clinicians, which complement and inform their research, whereas only a limited number of political scientists simultaneously work in the political or policy arena – though increasingly, many collaborate with "implementing partners." In turn, the resources associated with bio-medical research are exponentially larger than what is available for social science. While the findings from much bio-medical research are embargoed until published (in a timely manner) by a journal, most political science research has been discussed and widely circulated long before it is published in journal form (and very rarely in a timely manner).

Nonetheless, I believe that the bio-medical research cycle offers insights for social scientists that are worth considering, particularly for those who seek to make greater progress in gathering evidence that can contribute to knowledge accumulation and in some cases, practical knowledge that might have policy-related or other relevance. Of course, I recognize that not all political science research is advanced with such aims, nor with great optimism about knowledge accumulation, and the intended lessons of this essay simply do not apply to such work.

While the leading scholarly journals of any scientific discipline are not democratic reflections of the interests and priorities of all who participate in the field, they are intended to be the most important outlets for scholarly research. Publication in such journals is rewarded through likely impact on the field and individual professional promotion, and thus, what is published in such journals helps shape the research agenda of many in the field. Thus, it makes sense to focus one's attention on what is published in such journals.

Arguably, the most important scientific journals of clinical medicine are the *New England Journal of Medicine (NEJM), The Journal of the American Medical Association (JAMA)*, *The British Medical Journal* (BMJ), and *The Lancet*. These outlets provide a lens onto what is considered the most substantively significant research in that field.

For example, consider the July 30, 2015 issue of the *NEJM*. Not surprisingly, it includes an article on a large scale RCT: the authors report the findings from a study of the effect of hypothermia on organ donors. And the article demonstrates exactly why randomized studies are so valued: We would be extremely unsatisfied with a study that simply reported a correlation between hypothermia and health outcomes following transplants because we would always wonder, *what was it* that caused some doctors or hospitals to implement this protocol and not others? Deliberate randomization does a great job of addressing potential confounders (selection effects and omitted variables that might affect health outcomes), and at the moment, we lack a better strategy. The article reports that an interim analysis revealed that the differences in health outcomes were so profound that the experiment was "discontinued early owing to overwhelming efficacy." Treatment was associated with much better health outcomes and given the research design, it is quite reasonable for us to infer that those outcomes were *caused* by the fact that they received treatment.

Political scientists could understandably envy the clarity and significance of such results. Just imagine running a field experiment on foreign aid and development and finding that a deliberative discussion led by village elders helped to produce a 30 percent decrease in funds leakage, and the World Bank and USAID insisted that the experiment be terminated early because of the demonstrated efficacy of the intervention! Indeed, this is the type of solid scientific evidence that many modern social scientists aspire to produce, and I believe it fuels the enthusiasm around impact evaluation research.

For many of us who read only the big headline stories about new medical breakthroughs, and who speak frequently about treatment and control groups, it would be tempting to imagine that the leading bio-medical journals are themselves outlets for *just* this singular (experimental) type of research, and if we want to accumulate knowledge that this is all we should be doing. But in fact, if one reads on, one quickly sees that many other types of contributions reach the peak of this scientific community.

For example, in the same issue of the aforementioned article, we find purely *descriptive* articles. One epidemiological study (an important sub-field with few causal aspirations) reports on the incidence of pneumonia in two American cities; another reports on the characteristics of a previously undocumented cancer variant. In neither case do the authors advance any real causal claims, but they do provide a rich set of analyses of important outcomes of interest, using their scientific expertise to accurately describe these phenomena.

Also of note, the journal reports not simply "definitive" RCTs, but early-stage research findings. As is well known, bio-medical researchers classify "phase 1" studies as proof of concept exercises in which a new drug or regimen is tested in a very small group of people; and "phase 2" studies as experiments conducted with larger groups of people to test efficacy and to further evaluate safety before risking the cost and expense to many more people. “Phase 3” trials – large-sample experiments on patients -- are conducted only once prior research has demonstrated safety and efficacy. Because of stringent ethical rules around bio-medical experimental research, large-scale RCTs are frequently not possible without clearing preliminary hurdles. But the results of earlier studies are still considered important scientific contributions in their own right: The aforementioned *NEJM* issue included a phase 1 study of a new tumor treatment with 41 patients; and a randomized, double-blind study of 57 patients in a phase 2 trial of a drug to reduce triglyceride levels.

Finally, there was a very short article with an image of strawberry tongue in a child that provides a clinical observation. That article was just one paragraph long, and highlighted that this observable symptom was used to make a clinical diagnosis of Kawasaki’s disease.

What makes this disparate set of articles evidence that a research cycle is at work? And what is the relevance for political science? In this one issue of a leading medical journal, we learn about a range of very different discoveries from problem identification all the way to the test of an intervention that would modify the outcome of interest in a predicted manner. Each is novel, but with different levels of uncertainty about the nature of patterns and causal relationships. At one level, we read of the most basic description – presentation of a visual image to aid in a diagnosis or classification; to some correlations; to some tentative theories about the effects of a new experimental protocol; to a late-stage RCT. Each is deemed a sufficiently important advance in thinking about substantively important problems. While no one would claim that a very brief case report would provide any deep answers to broader scientific questions of interest, as Ankeny (2011: 254) argues, clinical studies published in journals have provided an important foundation for the development of diagnostic (conceptual) categories, and the formulation of key research questions and hypotheses. The most famous example is the 1981 publication of case reports that turned out to be observations of what is now known as the acquired immunodeficiency syndrome or AIDS (Ankeny 2011: 258), and indeed, case reports comprise an important share of the aforementioned journals.

Moreover, when reading the late-stage experimental study, we see that it makes reference to a retrospective study. And the very differentiation of phased trials implies a step-wise, but cumulative path towards discovery.

Turning to other clinical journals, such as *JAMA, BMJ*, or *The Lancet*, a similar pattern is clear. And this applies also to other high-impact multi-disciplinary journals such as *Nature* and *Science*. To be certain, not all bio-medical research journals are as eclectic in the types of research published; and of course, I have not established here the actual influence of scholarship on particular research efforts within a scientific community. Nonetheless, what is clear is that leading scientific outlets clearly publish across the research cycle, and causal research is frequently strongly rooted in prior published studies documenting important discoveries.

And while the focus of this essay is empirical research, it is worth highlighting that within the leading bio-medical journals, normatively-oriented scholars frequently play an important role in various steps of such cycles, by commenting on the implications of particular sets of research findings, or by highlighting the need to focus more on particular questions. For example, following the publication of research demonstrating the effectiveness of HPV vaccines, and the subsequent FDA approval of the vaccine, an ethicist from a school of public health published a brief analysis of the ethics and politics of compulsory HPV vaccination in the NEJM. This article sheds light on a range of important considerations that will strongly mediate how scientific discovery actually affects human health and well-being, but in ways that were surely not explicitly discussed in most or all of the earlier stages of the research cycle. Normative analyses routinely appear in leading medical journals through editorials and "perspectives" pieces, which help to address the ethical dimensions of research practice as well as clinical- and policy-related developments. In short, normative work is tightly linked to the empirical research cycle.

Before turning to political science, I don't want to leave the impression that the bio-medical research community is of one voice on issues of causal inference. I think it is true that bio-medical researchers are generally *extremely* hesitant to assign causal attribution to *any* observational study. Rather than dismissing such work, findings are published as "associational" relationships. And in turn, policy-makers and clinicians are made aware of such findings, but are repeatedly reminded to apply such findings with great caution as healthy skepticism about *causation* remains. That said, within certain circles of the bio-medical research community, one can find similar types of methodological debates as those we engage in political science. For example, one pair of nephrologists (Kovesdy and Kalantar-Zadeh 2012) has recently written of their frustrations with a research paradigm that does not allow for causal attribution except in the context of an RCT. They argue for greater appreciation of techniques such as an instrumental variable approach!

The (near) absence of a research cycle in contemporary political science

In practice, political science research already proceeds as a mix of inductive and deductive research. That is, scholars (often in their everyday clothing as civilian observers of the political world; sometimes as consultants or research partners; or as part of the early stages of their research) come to observe new phenomena that disrupt their view of the *status quo*, sometimes against prior theoretical expectations. In turn, scholars describe what happened, come to theorize about the causes or consequences of such phenomena, often through observation of patterns studied formally or informally, develop causal propositions, and provide evidence testing those propositions with various types of data and analyses.

And yet, *what is very distinct in political science from the bio-medical model I described above is that most of those steps are very rarely publicly recorded as distinct scientific enterprises.* In fact, increasingly, only the last set of studies – those that test causal relationships, especially using evidence from research designs that explicitly avoid threats to causal inference from confounders, and designed to accurately detect null relationships[3] – are the ones that get published in top political science journals.[4] Many of the other steps are described in a cursory manner and barely find their way into the appendices of published work. In other words, it appears that increasingly, the only types of contributions are those associated with the final stages of the research cycle.

For example, if one looks at the eight empirically-oriented articles of the May issue of the political science flagship journal, the *American Political Science Review*, all eight sought to provide a fully worked-out theory; most were explicitly testing causal models. Only one was an experiment, while six used statistical analyses to analyze large datasets; and one used case studies. But in virtually all of these articles, the authors largely say or imply that they are providing the best answer to a causal question with causal evidence. My point here is *not* that they

[3] That is, avoiding "Type II" errors, the *false* failure to reject the null hypothesis.

[4] *Perspectives on Politics* is somewhat unique in this respect, because it publishes a wide range of studies, including research articles that are not strictly concerned with estimates of causal effects.

were all quantitative or all experimental, because there was actually a bit of diversity on those dimensions.

The vast majority of political science articles at virtually all of the top journals, and the papers presented by ambitious graduate students in search of academic jobs, are increasingly of a single type: claim to provide a new theory, specify some hypotheses, test it with analysis of a large dataset, frequently associated with an experimental or quasi-experimental research design, and on occasion, explore with a few case studies. A great deal of this work is excellent and in many ways, has provided much more reliable knowledge than what was published in prior generations. More and more political scientists have turned towards design-based experimental and quasi-experimental research, and the bar for what should be trusted as causal evidence has certainly been raised. As a discipline, we have developed a heightened appreciation for the range of confounders that limit our ability to imply causal relationships even when presented with strong statistical associations. And in turn, more applied researchers have focused on implementing "well-identified" designs, lest they be challenged for over-claiming the fit between evidence and theory. Excitement over a range of new strategies for making causal inferences has implied greater attention to such work in leading political science journals and in the profession more generally. Clearly, these are largely positive developments.

But alongside this trend, Gerring (2012) documents the virtual disappearance of descriptive studies from the leading political science publication outlet and indeed, part of the problem is that scholars are not particularly interested in carrying out "mere" description.[5] Moreover, the unspoken presumption that the best work ought to be confirmatory, or a test of an *ex ante* specified hypothesis, rules out the honest publication of findings of surprise patterns. While increasing calls for the public registration of pre-analysis plans are aimed to keep professionals "honest" by limiting post-hoc findings being reported *as if* they were confirmatory, such efforts may inadvertently devalue the potential importance of strong and surprising inductive, accidental, or *post hoc* findings that shed light on big questions.

Moreover, the normative portion of the discipline – what is generally referred to within departments and the discipline as "Political Theory" – largely operates and publishes in isolation from its more empirically-oriented counterparts. While the topics of democracy, violence, public goods provision, identity formation, and the like do largely overlap, true integration within research cycles is largely absent. One rarely finds theorists citing or commenting on the latest empirical research; and one almost never finds empirical researchers discussing more contemporary normative research.

[5] Top journals do sometimes publish purely descriptive articles, but these works almost always make a significant methodological contribution as well as a substantive one.

And finally, the research that seems to be disappearing most quickly from the heights of the discipline are those studies that fall in between *pure* description and very strong causal inference. What bio-medical researchers would describe as correlational studies, such as retrospective cohort studies, are like kryptonite to aspiring young scholars who have a good sense of how such work will be judged – irrespective of the potential substantive or theoretical importance of some such studies. We provide very little space for “tentative” or “suggestive” findings, insisting that research ought to be definitive, or at the end of the research cycle.

In many ways, I share the view that the focus on improving the quality of causal inferences marks an important and positive development in the discipline for both quantitative and qualitative research. We should not go back to the times of interpreting any old significant regression coefficient as evidence of a causal effect. But it is also worth taking a step back to consider what it might mean for disciplinary practice and output if the only studies that are highly valued are the ones that can unambiguously demonstrate random assignment to treatment, allowing for more certain identification of causal effects. What are the implications for the types of questions that might (not) get asked? What does this imply about the efficient allocation of resources, and transparency in research? Are there lessons to be learned from the bio-medical paradigm described above?

Costs: The crowding out of discovery; Premature experimentation

If we are ultimately interested in causal questions and causal evidence, shouldn't we focus our attention on research that identifies causal effects? If as a discipline, we lack a large body of definitive scientific findings, shouldn't we play "catch up" by gatekeeping out the types of more tentative and ambiguous research that simply leads to endless debate about model specification, etc.?

In fact, I believe that there are several important costs in terms of the potential discoveries that are not incentivized because they are not appreciated, and the potential mis-allocation of our human and financial resources towards experimental and quasi-experimental research, not all of which is as promising as it could be. What we might call "late-stage" RCT's are (generally) extremely expensive in multiple ways: They often involve substantial burdens on human subjects in terms of time for participation and/or enumeration, they can be very expensive to administer from a data collection standpoint, and if there are ethical implications, these tend to be multiplied on a large scale, all because experimental analyses require analytic power, which for most social science experiments (which tend to have relatively small treatment effects) implies large sample sizes.

In the bio-medical sciences, owing to the very clear threat to human life and well-being of ill-conceived treatments and protocols, phased research is generally required for research with human subjects. As discussed above, early stage studies tend to be smaller in scale, and look more holistically at possible secondary and unanticipated interactive effects. For example, an adverse outcome within a subset of treated subjects would demand a retrospective analysis of differences, and an inductive, *post hoc* analysis of the predictors of heterogeneous treatment effects. Such exploratory study can be usefully carried out within the context of a smaller scale experiment such that the findings, if deemed relevant, can be implemented in the design of subsequent, larger-scale studies.

But political scientists generally lack the equivalent opportunities to publish phase I or phase II trials. At the very least, we lack a shared understanding of the role that such work might play in a larger research cycle. Nonetheless, most ambitious field experiments ought to begin with some degree of piloting and qualitative research (Glennerster 2013; Paluck 2010), including, for example, more open-ended focus group discussions and interviews with subjects. Owing to costs and uncertainties, such pilot experiments are, by definition, not at a scale that allow sufficient statistical power to reach definitive answers to causal questions. The question is, should such studies form part of the “official” research cycle, in the sense of being published? Or should they remain part of the internal analytic support that largely remains hidden until the “full” study is completed? I advocate the former. At the moment, political scientists might exercise the option of writing a blog entry about their findings, but this clearly winds up being a temporary, insiders’ outlet, and particularly for young scholars, provides little professional reward. The lack of a peer-reviewed outlet reflects the low value such findings are currently ascribed.

Absent any obvious outlet to publish such studies in political science, most political scientists will find little incentive to conduct such work or to take it as seriously as they should. Rather, they are more likely to "go big or go home," in pursuit of results that limbo their way under the conventional p=.05 level of statistical significance.

Even before conducting early-stage experimental research, good scientific practice would demand that we at least try to establish plausible connections between variables with existing and/or non-obtrusive data. And yet, in the leading political science journals, it is increasingly rare to find an observational analysis that simply reports important and robust associations unless the author has identified some "as if random" natural experiment and can use some type of naturally occurring discontinuity to infer causation. Now, of course, we would rather find an interesting natural experiment if the costs in terms of external validity are not too great. But sometimes, this is not possible, especially for analyses of large-scale, macro-level processes. And why are retrospective observational studies not still valuable if scholars are honest about what they can infer, demonstrating they have made the best attempts to answer their research questions with available evidence (or all evidence that could be reasonably gathered)? Shouldn't predictive patterns provide some initial confidence that a causal relationship *may* exist? Correlation does not mean causation… but it certainly can be suggestive of and an important piece of evidence in support of a causal relationship. If we consider again the bio-medical model, the first studies that found an association between smoking and lung cancer were hardly definitive, and we still would not run a randomized study to test the direct causal relationship. But the significance of the finding cannot be overestimated, particularly as scientists have concluded with mechanistic evidence (and without experimentation on human subjects), that smoking causes cancer.

And yet, for young social scientists – the ones most likely to be making new and creative discoveries, and perhaps the least well positioned to be raising vast sums of money for large-scale experiments – increasingly "causal identification strategy" is the only name of the game. And if they are not implementing proper experiments, they are seeking out causal research projects through "natural experiments" (Dunning 2012). That is, they search for the perfect "plausibly exogenous" instrumental variables such as rainfall or other arbitrary cutpoints and decision-making rules. And *de rigeur*, they are expected to proclaim that their particular strategy is "novel," and/or a rare "exploitation" of an untapped inferential resource.

To be sure, many such studies *are* exceptionally creative and valuable. And the sometimes quite clever identification of naturally occurring experiments is a feat that deserves proper accolades and professional rewards… But if the proverbial tail is wagging the dog – that is, if researchers wind up looking for outcomes to study because they finally stumbled upon an exogenous source of variation of “something” that ought to be consequential – well, that seems not to be the basis for a promising or coherent research agenda. There may be undue temptations for false discovery – i.e., “I’ve found something exogenous, now let me try to find *some* plausible outcome that it can predict,” in which case we may wind up with the same types of spurious associations that experimentalists have been trying to avoid. (I discuss the potential use of pre-analysis plans later in the essay.) Moreover, I think that many will agree that way too many recent social science papers are making overly-heroic claims that particular choices or events are plausible instruments and that they meet the necessary claims to make causal inferences.[6] I suspect that if our vision of good science explicitly allowed for “tentative,” “suggestive,” or “predictive” findings, we would see less over-claiming about the strength of causal evidence.

The increasing focus of talents, energies, and professional rewards on causal research *per se* poses several additional costs:

First, it likely obscures timely documentation of potentially important new descriptive discoveries, at least by political scientists, with the skills and insights

[6] For example, the exclusion restriction is rarely plausibly met in political science applications of instrumental variable analysis.

they could bring to such research. Such descriptive analysis ought to be both an end in itself, and also a gateway to other types of observational, and experimental studies (Gerring 2012). Along these lines, the discipline has turned away from a legacy of description at potentially great cost. Of course, it is not possible to account for the studies that might have been written and published had they been properly incentivized, but at the very least, we can say that much has happened in the political world in recent years… and political scientists have documented very, very little of it, at least in our leading journals!

Perceptions of disciplinary norms and expectations weigh heavily and are self-reinforcing. For example, today, if I had a graduate student who was working in rural Zimbabwe and identified some new form of interest articulation, or deliberation that we had never seen before… let's say that they had developed a pattern of decision-making in which they decided that all of the children in the village got to decide how to manage the local budget… I am fairly sure that the *only* way in which that graduate student could get that observation published in a top journal would be to figure out some way to observe tons of variation in the manifestation of that institution, and then to develop a theory of its causes or consequences, and then, test that theory by identifying natural randomness in the causal variable, or to run an experiment in which the institution itself was randomly assigned. And if that graduate student, who found this extremely interesting new aspect of political life that we had never seen before could not do all of these other things, I would need to, in good conscience with respect to her professional prospects, tell her to drop this immediately. Maybe she could publish in some obscure area-studies journal, but definitely not in a political science journal.

In a similar manner if a political scientist had been able to rapidly conduct a survey of social and political attitudes of Cubans just after the thaw in U.S.-Cuban relations, it strikes me that we would want to document such attitudes, to do it with the best social science skills available, even if the research had no ambitions of being able to detect specific causal effects. Whether Cubans favored political reform or not – the answer, particularly the distribution of responses, would be intrinsically interesting – and that piece of research could generate deeper inquiry about the causes or consequences of such sentiment. But in the near-term it would be a truly significant contribution to knowledge, simply as a piece of descriptive inference.

The point is not that political scientists should be reporting the news. They should be using their conceptual, analytical, and measurement skills to describe patterns and phenomena about contemporary and historical political life that would otherwise go unrecognized.

At the moment, however, apart from making valuable contributions to blogs such as the extremely popular and successful *Monkey Cage*, scholars are not incentivized to use their sophisticated tools to describe what is going on because again, we do not reward those contributions with our central currency: publication in peer-reviewed journals. Moreover, non peer-reviewed blogs are not intended for in-depth scholarly studies, and they do not provide an opportunity for disclosure of research methodologies, uncertainty of estimates, etc.

By contrast, in the bio-medical sciences, when a new set of life-threatening or otherwise critical symptoms present themselves, particularly in a patterned manner, one can be certain that such discoveries will be reported in a top journal with the expectation that *future research* will be needed to understand the causes and consequences of such discovery, and to develop interventions to prevent or to treat those symptoms. As discussed above, the *New England Journal of Medicine* published an article describing a Strawberry tongue, in effect communicating, “Hey, this is important, take a look. More later.”

I believe this state of affairs dis-incentivizes novel discovery, and incentivizes work within a narrow band of research in which processes and measures are already well understood. It is true that much of “normal science” involves small and marginal revisions and even just replications of prior studies. Such work deserves an important place in the discipline. But there also needs to be a place for examination of previously unexamined phenomena even if the causal connections are not fully worked out. In recent years, many graduate students – including those that have been extremely well trained in the best methods of causal inference – have confided in me that they feel “paralyzed” by the emphasis on causal identification and close out certain types of research questions very quickly because they don't believe that they will eventually be able to estimate a causal effect in the manner that they perceive the discipline now expects.

Towards a framework for a research cycle

If the concerns about the need for a publication and professional opportunity-incentivized research cycle in political science are valid, what is to be done? Importantly, I think we need to distinguish and to label the different types of contributions scholars might make, and to establish standards of excellence for each. (Although in the discussion above I identify an important place for normative research in the cycle, I do not include here a discussion of standards for such pieces. Normative contributions might be made at any stage of the cycle.) Not all journals will want to publish pieces from all stages; and the specific contributions of any piece are likely to be unique and subject to scholarly tastes and concerns. Nonetheless, authors, reviewers, editors, and readers should identify quite explicitly where in the research cycle any given study is likely to fit, and thus, how to evaluate the nature of the contribution. Our expectation should not be that every paper would tackle every concern within a substantive research agenda, but that it will take its proper place within a larger division of labor.

I follow Gerring's (2015) "criterial approach" and an appreciation of tradeoffs in research as a framework for making distinctions between types of studies (Table

1), but with a focus on the research cycle. A key tenet of good social science research is to avoid “over-claiming.” That is, do not attempt to draw conclusions that your data cannot support. But if we are going to provide a framework for honest research, we need a greater diversity of the types of claims that we might make, and associated standards for excellence and importance. What is critical about the notion of a research cycle is that we ought to value new contributions *based on what has come previously within that substantive area of research.* This, of course, places a particular burden on scholars and reviewers to be cognizant of what has and has not been learned in an area of research, and to properly frame contributions with respect to such background. While this might seem obvious, I think it is a point worth emphasizing in order to guard against the simple application of a single set of standards (i.e., what is the strength of the causal identification strategy?) to all scholarly work.

I describe below several broad types of studies, and contrast them in terms of the nature of the claims they make, and how they might be evaluated based on novelty of the descriptive or causal theories associated with the claims; the strength of association or effect size; the additional credibility associated with a publicly registered pre-analysis plan, and other considerations for evaluation. In all cases, high quality measurement of constructs is a pre-requisite for excellence: if constructs are not properly measured, no results can be considered trustworthy.

In each case, our criteria for a “significant” study should be to disrupt some aspect of prior knowledge. Critically, however, not all studies can or should contribute along every dimension.

A descriptive study in political science ought to use the best-available conceptual, measurement, and sampling tools to depict a phenomenon of interest. What are citizens’ political attitudes? How have certain institutions evolved over-time? In order to be considered important, such studies generally need to focus on a subject that is truly novel or that disrupts conventional wisdom about a particular state of affairs: for example, documenting either a new type of institution or set of political attitudes or behaviors; describing some aspect of political life in the wake of an important historical moment; or showing that a particular way of understanding some existing phenomenon is no longer correct, given superior data or measurement techniques, which in turn might cast some existing causal theories in doubt. These are akin to the bio-medical case studies or studies that simply describe the prevalence of a particular disease in different locations, perhaps reporting on associations, with no claims of estimates of causal relationships. The field of epidemiology provides critical insights for the bio-medical sciences more generally by offering careful description of the pathogenesis of disease. In a similar manner, political scientists could and should be making important and methodologically sophisticated contributions by describing the prevalence and variance of key political phenomena. And with the advent of “big data,” I expect that many such contributions will be advanced along these lines. In their seminal work, King et al. (1994) discuss descriptive inference at length, but that part of the methodological treatise is routinely ignored. An important descriptive study must demonstrate an outcome or pattern of interest that was not previously observed or expected, and such findings should open up new areas of inquiry within a research cycle. Descriptive studies may be retrospective (tapping existing observational data) or prospective (for example planned surveys). Fundamentally, these studies must be judged in terms of whether they demonstrate something that is truly new and if they are carefully measured/implemented.

Beyond description, analysis of observational data of naturally occurring phenomena can be used to detect patterns and the strength of relationships among variables. Within such studies, political scientists will make claims about the extent to which relationships might be interpreted as truly *causal*, providing not simply statistical or qualitative assessments of uncertainty in the strength of relationships, but additional discussions of the credibility of the research design, and the ability to address rival explanations. All studies, of course, face the “fundamental problem of causal inference,” (Rubin 1974) which is that we cannot know for sure what the counter-factual outcome would have been if particular units had received different values on the explanatory or treatment variable. Most non-experimental studies exhibit a set of hallmark limitations in this regard: we do not know for certain the process by which treatments were assigned and if the selection criteria were potentially biased in a manner that is correlated with the outcome of interest. Thus, the onus on retrospective studies trying to advance causal claims is to show that a wide range of other rival explanations are not driving the results. In turn, much scholarly attention focuses on the credibility of causal inference depending on the “identification strategy” or “identifying assumptions.”

Indeed, some research designs do, in practice, appear to provide more credible estimates of causal effects because they have found a way of leveraging some phenomenon that is “plausibly” random. Whereas for other studies, more questions remain at the conclusion of the study concerning whether the key treatment or explanatory variable was truly exogenously assigned, and given that uncertainty, it is difficult to conclude that any estimated relationship reflects a *causal* process. In table 1, I distinguish those studies that can credibly claim to be leveraging a true natural experiment from those that do not, labeling the latter “Associational/predictive” studies.

And here is the fundamental rub: if we cannot be convinced that the treatment variable is truly exogenous – if we are always left wondering whether some omitted variable has confounded the results – can we really believe that the research output is significant and worthy of publication at a top journal? Or the basis for professional recognition?

My answer is that strength of causal identification strategy should be considered as just one criterion among several. And again, this is where I think the notion of a research cycle sheds important light on how to evaluate a contribution: In the early stages of a research cycle, we might heavily weight the extent to which the estimated relationship between variables represents a novel and theoretically innovative association, and the extent to which the demonstrated strength of that relationship is substantively significant. Such associations might be demonstrated through careful model-based statistical analyses or (comparative) case studies.

By contrast, in the latter stages of a cycle, particularly if a strong predictive pattern has already been empirically demonstrated, we should hold studies to standards that more credibly detect *causal* relationships with less tolerance for potential confounding. Specifically, here we should expect research that does a better job of approximating a “natural experiment,” and we would expect to see, for example, regression discontinuity designs, effective use of instrumental variables, and/or difference-in-differences designs, which might more directly address the threat of confounders to causal inference as compared with a more straightforward regression or matching approach to analysis (for example, see Angrist and Pischke 2014). In an analogous manner, qualitative research at this stage in the research cycle would need to reach a very high bar of addressing potential confounders with explicit evidence. To the extent that researchers develop strong and credible causal research designs for testing well-motivated causal claims, we should be *less* concerned with the extent to which effect sizes are small or large *as a criteria for publication* or for professional merit more generally. We will need to depend on scholars to adequately frame the nature of the contribution and for expert evaluators to assess the particular contribution relative to prior work.

Moreover, at the earlier stages of the cycle, the correlational study or its qualitative analog ought to be theory-motivating. In turn, if observed correlations are weak, or if the case study research finds no clear pattern or logic, the contribution is ambiguous and almost certainly not worthy of publication. On the other hand, at the latter stages of a research cycle, when expectations about a theory are greater, a research design that more credibly isolates the effect of X on Y ought to contribute to knowledge irrespective of the actual findings. The better the test (i.e., the less likely the research design is to report a null result when a causal relationship actually exists), the less we should be concerned about the specific results as a criterion of scholarly review. Of course, substantively large findings will always be more likely to gain more attention, all else being equal, but that is a separate issue from scientific merit.

Finally, there are experimental studies in which assignment to treatment is randomized by the investigator. Building on the bio-medical paradigm, I propose that political scientists would be well served to distinguish between early-stage and late-stage experiments:

Early-stage experiments should be designed explicitly as way-stations for larger-scale, costlier experiments, particularly when little experimental research has been previously conducted in this research area. While social scientists are currently not expected to adhere to phased research standards akin to clinical trials, in many circumstances there would be great value to such practice. Because early-stage studies are, almost by definition, underpowered (there are not enough subjects or observations to confidently "fail to reject the null hypothesis"), the criteria for publication or contribution to knowledge should not be the magnitude or statistical significance of estimated effects. Rather, an article reporting on an early-stage experiment ought to provide deeper insights into the fit between treatment and real-world or theoretical constructs; discuss ethical implications of the experiment; highlight qualitatively observed processes that link (or impede) the relationship between treatment and outcome; and the specifics of an innovative experimental protocol. The criteria for publishing articles that document such studies is the extent to which the analyst provides strong evidence to motivate or to discourage large-scale experiments with the same or a related protocol. Through description of preliminary results, description of focus-group or other interviews, and other observations, such articles can more definitively assess the promise of carrying out potentially difficult and costly research, even if the estimates of causal effects are more tentative.

By contrast, late-stage experiments should be judged to a much greater extent in terms of the extent to which they provide unambiguous tests of the effects of X on Y. By definition, they should not be underpowered, which makes them uniquely suited for drawing conclusions about null relationships. But beyond that, experiments can be judged on the extent to which they are implemented in a manner that fully address potential confounders in as efficient a manner as possible. Articles reporting on large-scale, late-stage experiments should not be judged primarily on theoretical innovation or novelty of association: such novelty ought to be established in less costly ways, earlier in a research cycle. Instead, late-stage experiments ought to be clean and definitive tests of well-motivated hypotheses. If social scientists (and funders) were to take the notion of a research cycle seriously, they would not carry out expensive or potentially unethical experiments in the absence of one of the earlier studies providing strong suggestive evidence of the merits of the hypothesis under examination.

What role for registration / pre-analysis plans?

A welcome trend that has already been imported from the bio-medical sciences to the social sciences is the practice of public, pre-registration of design protocols and analysis plans prior to the fielding of experiments. In the bio-medical sciences, this has been an important corrective to the burying of null results and post-hoc finding of "positive" results obtained from "creative" re-analysis of data.

Although a full discussion of the merits of pre-analysis plans is beyond the scope of this essay, it is worth reflecting on their potential role within the context of a research cycle. The goal of pre-analysis plans is to keep scholars "honest" and to avoid “p-hacking” – the search for results that accord with conventional thresholds for statistical significance through *some* combination of variables in a post hoc manner, after predicted findings were not attained. This is a worthy goal, and for a great deal of research, I fully support the use of such plans.[7] Not only should such planning and public registration deter false discovery, but it ought to provide a public tool for justifying prospective research in the first place. As I argue in table 1, for late-stage RCT’s, such registration is critical and it is difficult to imagine a strong counter-argument against their use *for that type of research*. Even for early-stage RCT’s, scholars ought to pre-register their research designs and pre-analysis plans, but our criteria for the significance of the contribution of a paper should not be as closely tied to those plans as would be the case with a late-stage RCT.

A more difficult question concerns the value of pre-registration of retrospective and non-experimental studies. On the one hand, for observational research taking place *late* in a research cycle, pre-registration may indeed provide great value. If a scholar publicly registered that s/he was going to investigate a particular set of archives in a particular way, and predicted a set of patterns with

[7] See, for example, Humphreys et al (2013), and discussion and guidelines at the Evidence in Governance and Politics (EGAP) website: http://egap.org/content/registration.

a pre-specified analysis, of course, it would be very convincing and impressive to find those patterns observed in analyses conducted after the data were collected (assuming, of course, a logical theory, sensible data collection, and sound analysis…) All else equal, such a study would be more credible than one that did not pre-register hypotheses and analytic strategies.

But again, if we take the idea of pre-registration too far, particularly if we develop norms in which scholars perceive that their "hands are tied" to report only those analyses that have been pre-specified, we will surely crowd out the important inductive work (some call it fishing) upon which scientific discovery depends.[8] Let me return to the (bio-medical) example of HIV and AIDS. On the one hand, in the later stages of understanding this disease, science, and frankly, humanity, clearly benefits from scientific practice that insists on pre-registration of trials around the efficacy of drug treatment. We would not want the practitioner community to be confused about what actually *works* because the only studies available to them were the ones that demonstrated positive results. Drug trial registries help to solve this problem.

---

[8] Thoughtful advocates of pre-analysis plan registers have explained simply that we ought to simply make distinctions between analyses that were pre-registered and those that were not, but to feel free to report both.

On the other hand, let's consider the process of discovery around the important question of what causes the *transmission* of HIV? This research clearly involved lots of inductive pattern-detection, particularly in the early stages of the epidemic. I recognize that early recognition of the association between sexual orientation and AIDS symptoms generated some awful inductive theories (famously, the Rev. Jerry Falwell declared AIDS was a punishment from God), but also was a necessary pre-requisite for valid scientific discovery of the pathways for transmission. It is difficult to reasonably imagine that such relationships could have been predicted *ex ante*, or for that matter, hypotheses about the protective benefits of circumcision, but these have proven to be unimaginably consequential discoveries for curbing the epidemic. If gatekeepers in the bio-medical community had restricted such knowledge because the research designs were not "causally well-identified" or a pre-analysis plan was not on file, one can only imagine how many more lives would have been lost to the epidemic.

Recognizing that registration of studies is not a pre-requisite for all forms of important research in the bio-medical sciences, political science should avoid being overly restrictive and we should not necessarily value a study more than another on the sole criteria that one was pre-registered. To be more precise, the value of pre-registration depends on the type of study and place in the research cycle. In fact, because social and political phenomena are surely much less predictable and mutate more rapidly than bio-physical phenomena, I would argue that much less of our research ought to be constrained in this manner.

Specifically, as I outline in Table 1, I find only limited value for registration of studies other than prospective RCT's. Where scholars are able to pre-specify research plans with some confidence, by all means, they should do so. At the extreme, of course purposive research is better practice than "barefoot empiricism." But particularly at the early stages of a research cycle, we should not expect that scholars will know exactly what they are looking for before they have looked. (That said, they should not claim *ex post* that they knew what they were looking for when their findings were actually a surprise.) Problem-oriented research starts with puzzles about outcomes, and the search for plausible predictors of those outcomes is necessarily inductive. It is not always easy to judge whether findings from such studies are trivial/spurious or the advancement of real knowledge, but if other scientific programs are any guide, we should not restrict such inquiry wholesale.

For retrospective studies that advance a causal identification strategy involving a "natural experiment," public pre-registration plans could be a useful disciplining device, but their use should not give readers false confidence in the results should they be consistent with predictions. By definition, a retrospective study implies that the events of interest have already occurred, and it is often difficult to imagine that a researcher proposing to study causal relationships in a particular context will not have *some* prior knowledge of patterns in the data. As such, the finding of consistency between actual results and pre-registered hypotheses may not be as powerful as they appear. At the very least, pre-registration of analysis plans for observational data ought to welcome discussion of what has already been observed and analyzed.

Conclusions / recommendations

The notion of a research cycle as described here allows for the fact that intellectual progress requires many different types of contributions, and the quality of those contributions ought to be judged in terms of distinct criteria. Good research designs that allow for strong causal identification are critical for ultimately arriving at credible answers to causal questions, and these are most likely to generate knowledge that could be usable for advancing normatively attractive goals. Notwithstanding, well-executed descriptive or correlational studies also have very important roles to play in advancing such knowledge, particularly at early stages in a research cycle. Not all research questions are immediately amenable to the most definitive strategies for causal inference, but this alone should not be a barrier to pursuing substantively important research at the earlier, more tentative stages.

Good science should be public. It should be honest. And it should be cumulative. Right now, our structure of publication, reward, etc. does not provide the right incentives for all of these goals or a good division of labor in the form of a research cycle. Political scientists could collectively make greater contributions to knowledge if we built stronger scientific foundations with a greater diversity of research techniques and allowance for recognition of different types of claims.

How could research cycles, as described above, play a greater role in the discipline? The most important agents in this regard should be the editors and editorial boards of our leading scholarly journals. First, editors could more explicitly recognize a larger range of research contributions within their journals and label them as such, perhaps incorporating some of the language I have used above. Second, they could provide guidelines for reviewers concerning the appropriate criteria to use when reviewing articles with particular aims. Third, we must figure out ways to incentivize a more rapid timeline from submission to publication. It simply will not be possible to use scholarly journals as serious anchors for the accumulation of knowledge if it continues to take well over a year, sometimes longer, between submission and publication for successful pieces.

And beyond the journals, academic departments will need to make clear how they value different contributions in the research cycle as a basis for promotion and tenure. If younger scholars knew that they could advance their careers with different types of contributions, they would be more likely to focus on a wider set of concerns than an almost single-minded focus on strategies for causal identification. In fact, some of the self-monitoring that occurs within academic conferences and workshops might shift to *dissuasion* from premature experimentation on the grounds I have described above.

To be clear, my point here is *not* that political science should try to look just like bio-medicine. Rather, I think that there are some surprising lessons to learn that are worth considering. Academic disciplines evolve according to tastes and norms, and some appreciation of how other disciplines operate may widen our scholarly palates. At the moment, it certainly feels as if we could do a lot better in leveraging the collective research talents that exist throughout the discipline to answer serious questions about the political world.

References

Angrist, Joshua D., and Jörn-Steffen Pischke. 2014. *Mastering 'Metrics: The Path from Cause to Effect*. Princeton University Press.

Ankeny, Rachel A. 2011. "Using Cases to Establish Novel Diagnoses: Creating Generic Facts by Making Particular Facts Travel Together." In *How Well Do Facts Travel? The Dissemination of Reliable Knowledge*, New York: Cambridge University Press, 252–72.

Collier, David, Henry E Brady, and Jason Seawright. 2004. "Sources of Leverage in Causal Inference: Toward an Alternative View of Methodology." In *Rethinking Social Inquiry: Diverse Tools, Shared Standards*, eds. Henry E Brady and David Collier. Berkeley, CA: Rowman & Littlefield and Berkeley Public Policy Press, 229–66.

Gehlbach, Scott. 2015. "The Fallacy of Multiple Methods." *Comparative Politics Newsletter*.

Gerring, John. 2012. "Mere Description." *British Journal of Political Science* 42(04): 721–46.

Hainmueller, Jens. 2012. "Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies." *Political Analysis* 20(1): 25–46.

Humphreys, M., Sanchez de la Sierra, R., & van der Windt, P. (2013). Fishing, Commitment, and Communication: A Proposal for Comprehensive Nonbinding Research Registration. *Political Analysis* , *21* (1), 1–20.

Imai, Kosuke, Gary King, and Elizabeth a. Stuart. 2008. "Misunderstandings between Experimentalists and Observationalists about Causal Inference." *Journal of the Royal Statistical Society. Series A: Statistics in Society* 171(2): 481–502.

Keele, L. 2015. "The Statistics of Causal Inference: A View from Political Methodology." *Political Analysis*: 313–35. http://pan.oxfordjournals.org/cgi/doi/10.1093/pan/mpv007.

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

Kovesdy, Csaba P, and Kamyar Kalantar-Zadeh. 2012. “Observational Studies versus Randomized Controlled Trials: Avenues to Causal Inference in Nephrology.” *Advances in chronic kidney disease* 19(1): 11–18.

Munck, Gerardo L. 1998. “Canons of Research Design in Qualitative Analysis.” *Studies in Comparative International Development* 33(3): 18–45.

Paluck, L. 2010. “The Promising Integration of Qualitative Methods and Field Experiments.” *The Annals of the American Academy of Political and Social Science* 628(1): 59.

Rubin, Donald. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology, 66(5), 688-701.

Table 1: A Criterial Framework for Assessing Contributions in a Political Science Research Cycle

Study type		Claims / Strategies for making contribution	Importance of criteria for evaluation			
			Novelty of phenomenon / theory being studied within research cycle	**Strength of association; statistical significance**	**Quality of measurement?**	**Value of ex-ante public registration of propositions (i.e., pre-analysis plan)?**
Observational	**Descriptive**	To describe novel or unexpected phenomena, including variation within a population.	Critical	N/A	Critical	Very limited
	Associational/ predictive	To demonstrate a novel and robust pattern potentially consistent with a new or existing theoretical proposition.	More important	Critical	Critical	Very limited
	Natural experiment	To estimate a specific, predicted causal effect, using a naturally-occurring, but plausibly randomly-assigned treatment.	Less important	Important	Critical	Limited
Experimental	**Early-stage experiment**	To assess the plausibility of a specific causal effect and other possible (adverse) effects, using investigator randomization as identification strategy.	Less important	Less important	Critical	Necessary
	Late-stage experiment	To estimate a specific, predicted causal effect, using investigator randomization as identification strategy.	Least important	Least important	Critical	Critical

Response to Symposium Reviewers

Evan S. Lieberman, MIT

I am grateful to the editor of *Perspectives* for organizing this symposium and to this group of colleagues, for their thoughtful essays and reflections. The five authors have contributed several helpful elaborations and posed a few important challenges to my "provocation," to consider how the bio-medical research cycle might serve as a model for the production of political science research. I appreciate the opportunity to respond to a few of the key points they have made, and to invite further discussion.

Is bio-medical science a reasonable model for *political* science?

A first challenge, raised by McClure as well as by Elman and Burton, concerns the validity of the comparison between political science and the bio-medical sciences, particularly because of substantial differences in scale. For example, McClure highlights the vast size of the flagship *New England Journal of Medicine (NEJM)*, both in terms of staff and readership. And although she applauds their "long-standing record of editorial responsibility to both the medical profession and the public," in a manner consistent with what I describe in my essay, she asks whether we should temper our expectations in a far smaller discipline? Elman and Burton point out that the funding and administration of bio-medical sciences are structured in a manner and at a scale so distinct from political science, that one cannot help but wonder if those are not necessary conditions for success?

I certainly acknowledge that given such differences in the respective enterprises, one could not possibly argue, “it works for them, it will work for us.” Moreover, probably few bio-medical researchers would describe their own larger enterprise as entirely coherent, self-conscious, or fully successful! None of these caveats should prevent us from considering the apparent use of research cycles as a heuristic for re-thinking how *we* interact as a scientific community.

First, and this is really my key point: because clinicians and public health officials implement scientifically validated treatments and protocols every day, often with life and death consequences, bio-medical researchers really must care about causal effects and sound research designs for establishing causation. *Nonetheless*, that community *also* values a great deal of non-causal research. At the very least, this ought to make political scientists take pause and consider if we also should value explicitly non-causal research.

To respond more specifically to the possible “disanalogy,” I don’t see why disciplinary scale should be a limiting factor for the development of research cycles. Surely we have enough political scientists, even segmented into sub-fields, to engage in some division of labor? Perhaps one area that would require additional resources would be in the administration of our scholarly journals. Right now, our mostly “volunteer” workforce of peer reviewers and academic editors is highly constrained. In order to realize the vision I articulate in the original essay, we would require much more rapid turnaround in the editorial review process than is currently the norm. I don’t have an easy solution, but I also think this is a “fixable” problem.

Another concern is whether I am fairly characterizing the bio-medical research process? (At least the clinical research process, as McClure highlights.) Elman and Burton wonder whether my characterization of self-conscious cycles within the bio-medical sciences would be evident if I had probed more extensively. Here, McClure comes to the rescue, as she independently reviewed a few decades of NEJM editions, and concludes, "On Prof. Lieberman's view of editorial virtue, *NEJM* thus does well with regard to both the clinical research cycle and the biomedical research continuum." Indeed, it was my own reading of many editions of various bio-medical journals, as well as attendance at various bio-medical conferences that led me to the conclusions that prompted me to write the essay in the first place. Virtually every issue of the leading clinical journals I cite provides clear evidence of research findings at very different stages of what might be called a "cycle." While of course, many papers cross boundaries and carry out multiple functions, the point is that it is very easy to find these different types of discrete contributions.

Are cycles really *cyclical?*

Although my original essay emphasizes the role of research *cycles*, two of the response essays correctly highlight that in my discussion of potential best practices, I seem to emphasize more linear progressions and *end*points. Hibbing disagrees with my assertion that we might description as an "end in itself." Falletti challenges my hierarchical ordering, that research "*progresses* from pure description, through correlational analysis and natural experiments, to phased randomized controlled trials (RCTs)." She goes on to argue that experimental research may "feed into…

other stages of research." However, Falletti takes the opposite view of Hibbing and argues, "that the research process may productively end at the observational stage."

What are we to make of this exchange? Debates about whether descriptive research can satisfy our scholarly appetites are perhaps matters of taste, but I think there are also other issues at play. On the one hand, I agree with Hibbing that ultimately, we always want to understand the root causes of interesting descriptive findings. On the other hand, I suppose my intention, and here perhaps Falletti would be sympathetic, is that even if we do not have empirical evidence or even a well-grounded theory, to account for an "interesting" outcome, simply describing outcomes and patterns may still constitute an important and valuable research finding. We could argue further that such observations are generally interesting because they might *speak to* prior causal research by challenging such accounts. Even if there were not an immediately clear path forward for providing an explanation or causal account, certain types of descriptive findings elucidate the political world in which we live.

What role for qualitative and multi-method research?

Elman and Burton challenge my assertion that bio-medical journals appreciate exploratory and descriptive work by highlighting some notable rebuffs to qualitative submissions to bio-medical journals. But in all fairness, I never claimed that those journals valued qualitative research, and I would highlight that much exploratory and descriptive work is not qualitative! As they cited the *British Medical*

Journal, I immediately went to that journal's most recent edition, and the first article I read was a study of blood pressure in people with type 2 diabetes. The conclusion?

> Lower systolic blood pressure than currently recommended is **associated** with significantly lower risk of cardiovascular events in patients with type 2 diabetes. The **association** between low blood pressure and increased mortality **could be** due to concomitant disease rather than antihypertensive treatment."[1] [emphases mine]

This is my point exactly! The authors asked an important question, analyzed some relevant high quality datasets, did not pretend they could make strong causal claims, but provided some important directions for future research by accurately describing the observed patterns. The leading bio-medical journals are replete with such articles and I would describe this as *descriptive and exploratory*. They recognize that much of scientific progress is *tentative.* Of course, it is reasonable to fear that casual readers will interpret the findings as causal, but the authors and the editors have clearly done their job in explaining the nature of the finding.

Moreover, the leading clinical journals are filled with individual case reports and narrative accounts of various kinds (which I would take to be a type of qualitative research). I would acknowledge that there is surely a *hierarchy* in how different types of contributions within a leading journal are valued, as they should be: the very interesting but brief clinical report as well as the long-term double-blind clinical trial may get published in the same issue, but the latter will be viewed as a more substantial scientific effort. The relevant point for political science is that the bio-medical research community does not simply expect the *New York Times* health reporters to document all clinically interesting findings – rather, the academic journals determine the sets of findings that ought to appear in their pages, in turn setting the agenda for what journalists will subsequently cover in the news.

Units, effect sizes and causal complexity

Falletti raises other important points about the relationship between "causal complexity" and research practice, and challenges my attempts to find some equivalence in the value of research outputs irrespective of effect size. Rather, she argues that many null and small effect findings are the product of weak theorizing and/or research designs. She is certainly at least partially correct in that diagnosis. On the other hand, if we accept that social scientific inquiry is uncertain and much about the political world is random, we need to acknowledge that even some very well designed studies with promising theories will fail to reject the null hypothesis. We need to find a good place in the research cycle for publishing and valuing such studies, with appropriate criteria.

Building on the bio-medical analogy, Falletti offers the important example of increasingly individualized cancer treatment, illustrating that some researchers and doctors are using detailed knowledge and understanding of particular processes to formulate very individualized conclusions about likely treatment effects that can only be "tested" at the individual-level. My guess is that bio-medical researchers are far from unified in their assessments of this emerging practice. But I agree with Falletti that this is an interesting and important cautionary tale, especially as political scientists tend to estimate average treatment effects over large samples, and emerging best practices are weighing against *post hoc* estimation of heterogeneous effects in RCTs. The example serves as a reminder of the potential problems of labeling any particular method as an unqualified "gold standard" for attaining knowledge.

A way forward?

Let me also highlight a few of the excellent, productive suggestions made by the various authors. Elman and Burton suggest as a step forward more explicit journal recognition of different types of contributions before we can demonstrate the value of research cycles.[2] I agree. And to that, I would add that other gatekeepers in the discipline need to highlight and to value the different types of contributions scholars can make.

Hibbing offers the view that we might benefit from journals that contain excellent syntheses of what has come before.[3] I agree that awareness of and consensus about the state of knowledge in a research area is certainly critical for an effective research cycle.

I think this discussion comes at a pivotal moment for our discipline: information is flowing faster than ever before in history; we have many more tools for analyzing very large quantities of data; and along with all these developments, the nature of political life seems to be mutating quite rapidly, with enormous consequences for the human condition. Political scientists can and should play a big role in helping to understand these myriad changes. But we probably need to re-think some of our own institutions for producing knowledge if we hope to make important contributions and to take full advantage of our collective skills, knowledge, and talent pool.

I view this exchange over the possible relevance of the bio-medical research cycle as one of many possible discussions we might be having about whether the surprisingly rigid institutions we have built up to produce social scientific research are really the best ones for the current context in which we work.

---

[1] Adamsson Eryd, Samuel et al. 2016. "Blood Pressure and Complications in Individuals with Type 2 Diabetes and No Previous Cardiovascular Disease: National Population Based Cohort Study." *BMJ* 354. *doi: http://dx.doi.org/10.1136/bmj.i4070 (Published 04 August 2016)*

[2] McClure correctly points out that at the time of this writing, I do chair the committee charged with identifying an inaugural editorial team for a new online-only, open access APSA journal. As a matter of transparency, I should say that my essay was drafted prior to being asked to chair that committee, and I don't believe that those who appointed me were aware I was writing such an essay, so there is no explicit relationship here, though it is certainly likely that both outcomes are related to my own interests in methods and the production of knowledge. To be clear, the views I have expressed here are my own, and not intended in any way to communicate a vision of what that journal ought to be.

[3] A point also made to me by John Gerring, personal communication.