Publication, Publication

Gary King, *Harvard University*

Introduction

I show herein how to write a publishable paper by beginning with the replication of a published article. This strategy seems to work well for class projects in producing papers that ultimately get published, helping to professionalize students into the discipline, and teaching the scientific norms of the free exchange of academic information. I begin by briefly revisiting the prominent debate on replication our discipline had a decade ago and some of the progress made in data sharing since.

A decade ago this journal published a symposium on replication policies in political science. The symposium began with an article I wrote entitled "Replication, Replication," and was followed by opposing and supporting comments by 19 others (King, 1995). The debate over proper policies continued for a few years in subsequent issues of the journal and a variety of other public fora. Since then, many journals in political science have adopted some form of a data sharing or replication policy. Some strongly recommend or expect data sharing and some require it as a condition of publication. The editors of the major international relations journals have collectively written and committed themselves to a strong standard minimum replication policy (Gleditsch et al. 2003). Most important, numerous individual scholars now regularly share their data, produce replication data sets, put these data sets on their web sites, send them to the ICPSR and other archives, or distribute them on request to other scholars. Scholars sometimes worry about being "scooped," about maintaining the confidentiality of their respondents, or about being proven wrong, but since authors who make their data available are more than *twice* as cited and influential as those who do not (Gleditsch, Metelits, and Strand 2003), the strong trend toward data sharing in the discipline should not come as a surprise.

Gary King *is the David Florence Professor of Government at Harvard University, where he also serves as director of the Institute for Quantitative Social Science. His homepage can be found at http://gking.harvard.edu.*

The broader scientific community both collectively and in many other individual fields is also moving strongly in the direction of participating in or requiring some form of data sharing. Recipients of grants from the National Science Foundation and the National Institutes of Health now are required to make data available to other scholars upon publication or within a year of the termination of their grant. Replicating, and thus collectively and publicly validating, the integrity of our published work is often still more difficult than it should be, and some still oppose the whole idea, but our discipline has made substantial progress.[1]

The original replication debate in *PS* included discussions about student involvement, and indeed some departments now require students writing dissertations and senior theses to submit a replication data set that, after an optional embargo period, gets made public and is permanently archived. In the decade since "Replication, Replication," and also in the decade leading up to it, I have tried other ways to help students benefit from this trend. Chief among these has been an effort to professionalize my students by, among other things, giving them first-hand experience replicating published work and publishing their own. In particular, I require my students to write a "publishable" empirical paper for their class project based on their replication of an existing published article. Indeed, most of this paper is taken from a handout I have edited and re-edited over 20 years to maximize the chance that students are able to publish the paper they write for a methods class I teach.[2] Students are told that successful projects need not actually be published or even submitted for publication. However, although writing a publishable paper may sound hard at first, revised versions of a large proportion of student papers every year eventually result in published journal articles, and many have also appeared as convention papers, dissertations, or senior theses, and they have won many awards. Almost all of those who closely follow the suggestions below wind up with published articles. The advice offered here is not the only way to conduct high quality research, but it is one relatively high probability path to success.

Some students ask: "Why begin an original research paper by replicating some old work?" A paper that is publishable is one that by definition advances knowledge. If you start by replicating an existing work, then you are right at the cutting edge of the field. If you can then improve any *one* aspect of the research that makes a substantive difference and is defensible, you have a publishable paper. If instead you begin a project from scratch without replication, you need to defend every coding decision, every hypothesis, every data source, every method—everything. In contrast, if you start with replication, you only need to defend the one area you are improving, and you can stipulate to the rest. If a critic doesn't like something else in the original article other than that which you are improving, you need not defend that point since it is already part of the published record and is the recognized state of the art. After all, this strategy was not originally designed for students; it is exactly the procedure followed by many faculty in political science and most other scientific fields. It is one of the reasons that the process of providing access to the raw materials of research with sufficient precision necessary to replicate, and of accessing that from other scholars, has become a deeply established part of the scientific process.[3]

What follows is some of the advice I give my students.

Elements of the Paper

1. Your paper should address a substantive problem in your field of interest and contain one or a few clear points; one point with several supporting points is better than a lot of unrelated points. Your point should unambiguously answer the question: *Whose mind are you going to change about what?* If that question isn't answered, then you're not making a contribution and there's little reason for the paper to be published.

2. Begin by locating an article in your field, acquiring the data used in the article, and replicating the specific numerical results in the tables and/or figures in that analysis. (You may start with the original article and find the data used, or work backwards from the data, such as stored in the ICPSR's Publication Related Archive, or one of the other

https://doi.org/10.1017/S1049096506060252 Published online by Cambridge University Press

archives of datasets constructed for the purpose of facilitating replication, and find the scholarly article.) This article should have been published in a peer-reviewed scholarly journal, preferably within the last 4–5 years, the more recent and prominent the better. The better the article, journal, and author you choose, and the more often the article has been cited, the more likely your paper will be publishable. Checking citation indices (ISI or Google Scholar, for example) is often a good idea, but be mindful of the selection problem that occurs because more recent articles will have had less time to be cited.

Please beware: replicating an article, even if you secure access to the original data, is normally a highly uncertain and difficult process. Analyses that look neat and clean in published articles often prove to be far from that in reality. Most students find that prominent articles by leading scholars in the field contain errors, confusions, lack of essential information about how the analysis was conducted, and other problems. Some of these issues do not matter to substantive conclusions, and some do, but all make replication more difficult. As such, completing the replication will likely be more troublesome and time consuming than you anticipate (even after you adjust for the information in this sentence!). After you have done everything you can do on your own, you may need to contact the author of the article (please do so respectfully and diplomatically).

(The remarkable difficulties students have in replicating published articles teaches more about the state of the literature, and conveys more about the sometimes shaky foundations of academic knowledge, than reading all the published literature one person could possibly consume on his or her own. Every year some students are incredulous or stunned by what they find; the experience is in part disheartening, but it also seems to empower students who (correctly) conclude that they can do better. I typically devote some time during class to share these experiences.)

3. Please bring me a copy of the article you choose and ask for my views before proceeding. This will generate advice on what is unlikely to work, and might be useful for other reasons, but to be clear it is no guarantee that you will be able to replicate the work chosen and successfully complete the assignment. Your assignment is to pick an article according to the criteria above *and* to replicate it. The choice of the article is part of the assignment and so, just as happens to faculty researchers, you may need to change your choice of topic along the way depending on what you find or difficulties in replication and do it all again. Perhaps this is why they call it research, rather than merely search! (If you change articles, please bring the new article to me as well.) You may wish to follow the procedure that many of us follow by starting several projects at once and then following up those that seem most productive.

4. If you decide that the conclusions of the original article are incorrect, then show why you think that but also what led the authors of the original article to think otherwise. You should never discuss it in the paper—directly or indirectly—but you should assume, unless you have overwhelming evidence to the contrary and maybe even then, that the authors were well-intentioned, smart, honest, and hard-working. Your article is about the author's findings, not about the author.

5. Clarify with precision the extent to which you were able to replicate the author's results. If you can't replicate the author's results even with the help of the author that is important information that needs to be on the public record, but it also means you can't build on this work to make further progress. And if you can't find out what the problem is, it might mean that you do not have a publishable paper and so might need to start with a different article. So try hard, and you may have to try very hard, to replicate.

6. Unlike almost all previous papers you may have written, do *not* allocate space in your paper in proportion to how much work you put in accomplishing each task. The point of this paper is to make your scholarly point, not to show how smart you are. This paper should not be about you or a report of what you did; it should be about what you contribute to our collective knowledge about the world. For example, a large fraction of your effort will probably go into replicating a prior result (and thus getting up to the cutting edge of the field), but only in rare cases will that take more than a page or two of your paper. Space in your paper should be allocated in proportion to how much of a contribution it makes to changing the minds of someone in the literature about something important. Thus, for example, if at the end of week of data analysis you make one important finding on last day that would add to or change the conventional wisdom about a subject, then you should change the title, subject, abstract, introduction, and organization of your paper to focus on this finding. All your other efforts that, despite your best efforts, led to dead ends should be excluded from your paper unless they help you demonstrate this one key point. Resist the temptation to include all this just to demonstrate how much work you did; that's not the criterion on which you will be judged in this class (or afterwards). This task is a crucial aspect of your socialization into the profession, and your success requires that you learn it at some point. It might as well be now.

7. After replicating the article, follow the logic of King, Tomz, and Wittenberg (2000) and try to improve the presentation of the original results. See whether you can find useful, additional, or even contradictory information not discussed in the article without changing *any* assumptions in the original paper. If you are able to do this, then you need not defend anything other than your method of presentation, which would put you on very strong grounds in your claim for journal space. You may find Zelig (Imai, King, and Lau 2004) or Clarify (Tomz, Wittenberg, and King 2003) software helpful in calculating new quantities of interest from the same model.

8. Next, you should run some *controlled* methodological experiments designed to advance the state of knowledge about the substantive project. That is, make *one* improvement, or the smallest number of improvements possible to produce new results, and show the results so that we can attribute specific changes in substantive conclusions to particular methodological changes. (Improvements can include changing the way the author dealt with missing data, selection bias, omitted variable bias, the model specification, differential item functioning, the functional form, etc., adding control variables or better measures, extending the time series and conducting out-of-sample tests, applying a better statistical model, etc.) If you are able to produce an interesting substantive result that is different from the original article, with only one completely justifiable methodological change, then you only need to defend this change fully and carefully.

9. If you are able to improve or change the author's results in some important way with the minimal change necessary (and that is maximally justifiable), write that up separately. Then, in a separate section, go ahead and make all the changes you think are desirable and see what difference that makes to your results. But make sure the minimal changes necessary to produce the new conclusions are described and justified first with results fully presented. Once you've done that, then you're home free in your quest for journal space. The rest are further improvements that you will have much more free reign to explore as

https://doi.org/10.1017/S1049096506060252 Published online by Cambridge University Press

you see fit. But if it turns out that all these other changes don't change any substantive conclusions, then leave them out or report on them very briefly.

10. Provide evidence that your model fits the data in and out of sample (or perhaps that it fits better than the model in the original article). For example, are the probabilistic assumptions implied by your model correct? Do 95% of the data points fall within the 95% confidence intervals? Are you able to predict a set-aside test set from the rest of the data with the predicted level of uncertainty? You can test this for regression models, logistic models, and all other models. The basic idea is the same as any scientific analysis: look for as many observable implications of your theory as possible and to check those (King, Keohane, and Verba 1994).

11. Understand your raw data prior to statistical modeling, and help your readers do so. Include graphics or descriptive statistics to help in this goal. Giving some concise sense of the data while you are describing the variables is a useful space saving device.

Ground Rules

1. The paper must be coauthored with another member of this class.

(a) Why? Although papers are rarely coauthored in school, almost half of all political science articles are, which is a seven-fold increase since the 1950s (Fisher et al. 1998). This class is about research as it is actually practiced in the field.

(b) What if your coauthor doesn't carry his or her weight? Deal with it somehow, and make your best individual effort even if it is asymmetric. You will have to deal with this when you graduate too. Your goal (and given task) is to make your paper as good as possible, and you have at your disposal your effort and whatever effort you can marshal from your coauthor. In most of the social sciences, credit is not divided among the coauthors: each coauthor gets almost full credit for the entire paper.[4] As long as you're getting credit for what you're doing, it doesn't hurt you for someone else to have more credit than he or she deserves.

(c) But, some might scream, "it's not fair to share credit equally!" With all the time and mental energy you could devote to developing normative standards to apply to your collaborators, you could write another article. That would be a lot better for you and your career, and it will have the side benefit for the rest of us of making you a lot more fun to hang out with. It's also not fair that some came to this class with better math skills, or get to ski more often than I do, but such is life. A normative standard that is much more in your career interest is to ask yourself instead only: Is your coauthor making a *positive* contribution to your paper? If it's a positive contribution, then you're getting something out of your collaboration. Be thankful.

2. The authors on your paper should be listed alphabetically, which (in most social sciences) conveys equal contributions, or that everyone was a full-fledged member of the research team. Customs in public health and medicine, and in some other areas, usually give most credit to the first-named author, but it is possible even in those fields to convey to readers that contributions were equal, such as via explanatory footnotes (e.g., "authors were listed alphabetically"). Get these issues out of the way quickly so they don't affect your work.

3. Papers should be no longer than about 20 pages (double-spaced, one-inch margins, 12pt, including figures, tables, and references). Think in terms of a short research note, not a full-length article. Journal space is scarce and so the longer the paper you write, the harder it will be to publish. If you can do it in 10 pages, so much the better.

4. In addition to coauthoring your paper, get others to give you written comments on a draft version. Why? The reason academics hang out together in universities is not (necessarily) because we like each other; it's because our work gets better in the process of interacting. When you graduate, you will need to build a network of people who will carefully read your work before you distribute it widely; students in this class often form the start of that network for each other. But it is an implicit quid pro quo: If you want others to read your work, make sure to give them detailed comments too.

5. We provide a formal way to provide you some advice along the way: In class, you will turn in a very early draft of your paper with the tables and figures in near final form but relatively little text. You'll also turn in a replication data set, just as faculty routinely do. We will then give this to another student, who will try to replicate your results (without talking with you). That student will then write a memo to you about your paper, with copy to me. In science, we compete to advance knowledge about the world, not to tear each other down. Thus, the purpose of that memo is not to destroy anyone's work, but to improve it.

6. Do not ask the author of the published article whose work you are replicating for comments on your paper, and do not share it with him or her, or anyone outside of this class, until I have read it and you have revised it accordingly. Why? In all likelihood this experience will be your first interaction with the outside world as a professional academic and, like all academics, as a certain kind of public figure. The academic world has highly specific, and often unstated, expectations about a whole host of matters you may not now perceive, and authors tend to be very sensitive about what you write and how you write it, especially if you find something even slightly wrong with what they did. You can avoid a lot of trouble with a quick reality check. So please come by first.

7. After the paper is revised (for substance and style) to my satisfaction and yours, it will be much safer for you to go public, and going public then is essential. The procedure is, before you show it to anyone else outside this class, send a copy to the author of the work you're replicating or critiquing and respectively request comments. When you receive a response, you should revise, being as generous as possible, but only as you think is appropriate. Only at that point should you post the paper on your web site and make it fully public, which you certainly should do. If your contribution still stands, in your view, after receiving comments from a wider audience, you should then consider submitting the paper to a scholarly journal or presenting it at a conference. For information about where to submit your paper and how to do it, come by and we'll talk about it.

Style

1. Your paper should be rigorously structured and organized into sections and subsections. (Heading titles should be clear, contain no acronyms, and should summarize the key point you are making in the section. They may be numbered, but the numbers should be in addition to a substantive title.) The best way to understand how to organize a paper is to imagine that your readers will randomly fall asleep at any time for five minutes and yet keep turning pages; when they wake up, they should know exactly where they are from your subheadings alone. Something like this will surely happen: Think about what you do when you read long, boring textbooks.

You are writing for *anonymous* referees. Referees are busy people looking

https://doi.org/10.1017/S1049096506060252 Published online by Cambridge University Press

for a way to finish the thankless (anonymous!) task of reviewing your paper as quickly as possible. Since you're not likely to have as much time from them as you think, you need to make reading your paper as easy as possible. And in this game a tie doesn't go to the runner: If a referee didn't read carefully, pay attention, or understand you, or missed or misunderstood something important in your paper, it is your fault. And since it is your paper and not you that matters, anonymous referees will not (and for the sake of the literature normally should not) give an anonymous paper writer the benefit of the doubt. Anonymous referees are not normally prone to spontaneous generosity and do not generally impute favorable motives to authors who are not clear or impute appropriate assumptions when you leave them unstated. (And this business is no worse than any other: Human beings acting anonymously in other circumstances tend not to be especially nice either. You may have noticed that cars, which have drivers' identities mostly obscured, cut each other off all the time, but this almost never happens walking down the same streets.)

2. The overall structure of the paper, and all the key points you want to make, should make sense in terms of accomplishing *your* goal. If you include many section breaks it is easier for your readers to skip over things they are not interested in while still getting your point; if you include too few, they will get lost, and so will your chances at publication.

3. While writing, keep revising the list of section headings until it looks like a table of contents that conveys your key point well even if one does not read the paper.

4. Do not try to hide weaknesses in your paper. If you know of a problem with your analysis that you have not solved, clearly delineate the problem. If you think the problem is not that bad, explain why, but do so honestly. If you have an idea of how to solve it, but haven't done so, offer it as a suggestion for future researchers. If you don't know how to solve it, suggest that future researchers try to tackle it.

Why do you need to be so forthright with potential problems? If a reader sees a problem you didn't mention, you're making it possible for them to say "not only didn't the author correct this problem, but he or she didn't even realize it was a problem!" If all you do is to note the problem, you can take the edge off the criticism. This is of course as it should be, since your paper will also be a more appropriate scientific statement of the problem.

5. **Front matter**

(a) Your title should convey *your* key point by summarizing clearly your argument or angle. An appropriate title is not a list of topics or "the effect of A on B." Quite like haikus (which I recommend reading and writing for practice), writing titles takes considerable time, effort, attention, and thought.

(b) Include a footnote on the title page to the title, and put the text of the footnote at the bottom of the first page. In it, put your contact information, where others can get your replication data set, and acknowledgments to everyone who helped you with this paper. There is no cost to being generous with acknowledgments. Be sure to thank everyone who read the paper for you, including students who read it for a class assignment, anyone who you discussed it with, those who helped you solve computer or methodological problems, or anyone who provided you data. If you had any contact with the authors of the article you're replicating, be sure to acknowledge them too.

(c) Include a one-paragraph abstract, no longer than 150 words, on the page following the title page. The purpose of the title is to convey your entire point in one phrase, and to convince people to read the abstract. The point of the abstract is to drive home to readers your main contribution and "who's mind you're going to change about what." It should contain all relevant information about the importance of your work and who should read it, and not much else. Reading it should make you want to read the paper. By at least three weeks before the paper is due, send a copy of the title and this abstract to the class mailing list to get comments from me and others. You may send more than one version, and I may ask for revisions. This step improves the paper substantially by helping you to clarify the paper's primary contribution and to focus the paper on it. Once the title and the abstract are set, the entire rest of the paper is likely to be affected. Be sure to read all the abstracts and comments; this process is often extremely informative about how to write papers, about what is important, and about how to make the findings in your paper important.

6. **Appearance**

(a) Prepare this paper as if it were to be submitted for formal review at a professional journal. Why? Quality may be everything, but it is hard to measure and so style provides important signals. For example, as a purely predictive matter, papers formatted with LaTeX are much less likely to contain egregious methodological flaws. Similarly, you should run the paper through a ssppeellllliinngg checker. Use the appearance of your paper to your advantage.

(b) Follow the same rules used for preparing convention papers: Use 12pt double-spaced black text on white 8.5 × 11 inch paper with a staple in the upper left corner—no polywhatever colored plastic covers. The paper should have a title, your name, affiliation, and the class number. For examples, see the preprints at http://GKing.Harvard.edu/preprints.shtml.

(c) Follow this style for references "(Beck 1985)," etc., of the *American Political Science Review*. (Why? This is the most common style in the discipline and is quite common in other disciplines as well.) If you want details, see the instructions to contributors in the journal you are writing for, but most of these instructions should be ignored until acceptance since they are designed to make things easier for copyeditors, not reviewers.

(d) Avoid gratuitous citations to your professors (or anyone else). You are welcome to tell others how wonderful they are in person, but keep this out of your papers. Cite only those whose research you use or build on in some way. (Do not leave out those likely to be your reviewers either; no one likes to be ignored.)

(e) The paper should be a formal presentation, not a personal letter. Occasional humor is fine, but inside jokes or questions are best left out. Raw computer output should not be included.

7. In the text, identify the specific empirical question you are interested in immediately and get to it. Beginnings such as "In this paper, we demonstrate that . . ." are favored. If it is not clear to a reader what you are planning to accomplish with some specificity (including what your dependent variable is) after a few pages, something is wrong. As someone once said, if the first bite of an apple tastes bad, you don't keep taking bites to see whether some other part of it might be better.

8. In almost all cases, do not include a section titled "literature review," and any literature review you include should be

https://doi.org/10.1017/S1049096506060252 Published online by Cambridge University Press

short and directed toward your point only. Other people don't deserve to be cited in your paper unless they help you make your point; they already have their own papers. If prior literature doesn't help you make your key point, omit it.

9. If you have long technical lists of coding rules, or anything else that seems essential but distracts from your point, put it in an appendix.

10. **Be nice**

(a) Treat authors you are replicating as you would want to be treated. Your goal is to stand on the shoulders of the scholars whose article you are replicating, not to step on their faces. In all likelihood, there is a good reason why these people did whatever they did. Remember that if all goes well, you will one day be in their shoes.

(b) Talk about the *article* you are replicating, not about the *authors*. So you should write "Jones and Smith (2003) is mistaken" rather than "Jones and Smith are mistaken."

(c) Do not be personal and be careful of the language you use to describe your work. Remember that it doesn't matter whether you "agree" with the author you are replicating; no one cares what you "think"; no one is interested in your "opinion"; and readers don't want to know what you "believe." In your paper, *you* do not matter; the scientific community only cares what you can *demonstrate*.

11. Use active ("We ran a least squares regression.") rather than passive ("A least squares regression was run.") tense. Why? This is not only good grammar and easier to understand; it is a matter of standing behind what you're doing and sounding like you mean it.

12. **Math**

(a) Don't let your word processor control the look, style, or content of your paper. For example, mathematics in English text is, by scholarly convention, always in italics, and Greek letters should be in Greek. Math should not be written as yhat =beta-b + gamma-w *X +e, but rather $\hat{y}_i = \beta^b + \gamma^w X_i + \epsilon_i$. Larger equations should be set with equation numbers, and be referred to as with this example of Bayes Theorem,

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)}, \qquad (1)$$

where θ is an unknown parameter and y is a data vector.

Many word processors can do this (although the defaults often look lame); for a suggestion, I recommend LaTeX, which is a standard in most mathematically oriented fields, is used by many publishers, is free, and is available for most operating systems. Startup costs are higher and so it is not for everyone, but once you know how to use it is a lot faster than MS Word or other WYSIWYG programs. Introductory material can be found at my homepage.

(b) Equations included in papers should not be treated like figures. They are part of the sentence structure and so should include punctuation, etc. The equal sign is the verb in math, and so we say that $a^2 + b^2 = c^2$ (and read it as "a squared plus b squared equals c squared."). Or we explain that 5 is the result of evaluating 7-2. (Note also the periods at the end of the previous two sentences and the punctuation at the end of Equation 1.)

(c) For numbers, use only as many decimal places as you have precision. Normally one or two digits to the right of the decimal point is enough, but the right number depends on the context. Why? Think about the advice you would give to the local weather forecaster who predicts that tomorrow's temperature will be 37.828280019277647381 degrees Fahrenheit. Your standard errors provide a guide: If you present more digits after your decimal point than your standard errors indicate you can measure accurately, then you are filling your paper with numbers created by rounding error. (This silliness is not uncommon: if standard errors indicate that coefficients are accurate to 2 digits and 4 digits of accuracy are presented, then a majority of the numerals printed in the table are totally irrelevant.)

(d) A sophisticated reader must be able to write down your statistical model and likelihood function (or other method of estimation and analysis) from reading your text. You can convey this by writing down the precise form of your model, but do not derive or reproduce equations that are well known unless you cannot otherwise explain precisely what you did. For example, saying that you used even something as simple as a negative binomial regression model would require clarification since the variance function can differ across software implementations.

13. **Tables and Figures**

(a) Tables and figures should be included to make specific points, and to draw readers' attention to these points. They are not included to demonstrate that you did something. They are not obligatory every time you run a regression, for example. Readers will interpret your including a table or figure about a point as your judging it important. So choose carefully what you want to display in this way.

(b) All tables and figures should be separately and fully documented. Someone reading only them, without the paper, should be able to understand what is going on. Adding an explanatory paragraph at the bottom of each figure or table is usually necessary to accomplish this. Similarly, someone who reads the paper and ignores the table or figure should also be able to follow it all. The point of the text is to walk the reader by the hand through the table or figure so it is easy to understand. Picking out one number in the table and explaining it in detail at the outset as an example is often a good strategy.

(c) Do not add lines between every column and row in your tables (as is the usual default in programs like Excel and Word), and do not include superfluous shading of various sorts. If you have a column of percentages or dollars, only the item in the first row should have a % or $ sign. In columns of numbers, the decimal points should align vertically, and do not use proportionally spaced fonts.

(d) Make tables and figures only as large as they need to be; in most cases, small is beautiful so long as they can be read (including by people over 50). Remember that journal space is valuable. Try to keep tables and figures oriented as the text is (profile rather than landscape) so the reader doesn't have to keep turning the paper around. It is much easier for the reader if the tables and figures are run into the text (floated to the top of the page) rather than collected at the end of the paper or on separate pages or broken between pages.

(e) In most cases, if you can present the same information either way, a figure is better than a table.

(f) As in the journals, number the figures consecutively, and separately number the tables consecutively. Refer to each in the text by number (e.g., see Figure 4). The total number of tables and figures in your paper should probably be a single digit number.

https://doi.org/10.1017/S1049096506060252 Published online by Cambridge University Press

(g) Follow a no dumping rule. That is, each table and figure should be presented and discussed in the text in turn. You should not casually refer to Figures 1–12 and then go on to the next subject. Explain in detail what you want readers to see in them. Remember that what is obvious to you after looking at these for weeks will probably not be obvious to anyone else. Similarly, if you find yourself preparing big tables with lots of numbers and only talking about a few, then you should rethink your strategy.

(h) When explaining the content of a figure, it is good practice to devote one paragraph to the setup—the horizontal and vertical axis measurements, the unit of analysis, etc.—and then to start a new paragraph that explains your results.

Problems to Avoid and Other Suggestions

1. Decisions about what to present should be made by you, not by computer programs. You have been given the tools in this class to create your own statistical models, perform your own simulations and to calculate a quantity of interest from any model, if need be. The fact that R, Zelig, Clarify, or some other program does not do what you want is not a reason not to do it.

2. Don't do things like this:

- **Hypothesis 1**: The effect of . . .
- **Hypothesis 2**: Instead, the effect of . . .

It looks very scientific, but you want your points emphasized, not words in everyone else's paper (like "hypothesis"). If you want to emphasize something, *emphasize your point this way*. Numbering hypotheses and using scientific sounding words also doesn't usually help you make your point.

3. **Quantities of Interest**

(a) If you run some analysis, don't present long lists of coefficients that are (or anything else that is) hard to interpret. Instead, compute the precise quantity of interest (and a measure of uncertainty) that best helps you make your substantive point. (If you feel you must add the lists of coefficients to the paper, add them as appendices so they can be skipped easily.) No one cares about numbers that even the author doesn't want to interpret, and so these should not waste space in your paper.

(b) Don't write a paper that has a long buildup to an estimation and then have one table and a paragraph that summarizes all your work. Spend time *carefully interpreting* your results in substantive terms, in terms that a non-quantitative political scientist would understand. The goal ought to be to satisfy someone quantitative (by doing the statistics right) and a smart non-quantitative type (by fully explaining things in sufficient detail) in the same paper.

(c) Make sure all point estimates come with some measure of their uncertainty, such as confidence intervals, posterior distributions, or standard errors.

(d) Do not say that quantities are "statistically significant" unless you have a very good substantive reason to do so (hint: you probably don't!). In most cases, this is unhelpful information that distracts from the substantive purpose at hand. Calculate your quantity of interest by giving a posterior density, confidence interval, or point estimate and some measure of uncertainty. Once you've presented all that, you have conveyed all that your data have to say about your quantity of interest; what more would you want to know?

4. Don't claim to be using "the maximum likelihood model." ML is a method of inference. You talk about your statistical model, and then say you used ML to estimate it (if you did). Regression is ML, but it is more commonly understood as regression.

5. Don't include control variables in a model that are consequences of the causal effect you are trying to estimate (which is known as post-treatment bias). This is an important point that is often missed. To estimate two causal effects usually requires estimating separate models for each; although it may be possible to include both variables in the same regression, the coefficients cannot be interpreted as causal effects unless you are careful about this point. See King (1991) and King and Zeng (2006) on this point.

6. Examples: A full-length replication can be found at King and Laver (1993). Other shorter examples can be found in King, Tomz, and Wittenberg (2000) and King et al. (2001).

7. Provide sufficient information about your analysis so that it is possible for someone who reads your paper to replicate the analysis. This means that you must be very precise about coding rules, where the data came from, how indices were computed, what the unit of analysis is, etc. A class exercise will include another student replicating a draft of your work, but be sure the final version can be replicated too. Likely the only way you will be able to continue to revise the paper for publication after class is over is to prepare a replication data set now, while the work is fresh in your mind and all the data, files, and code you used are still available. You now know how hard it has been to replicate someone else's work; don't make the same mistakes.

If you reach the stage of publishing this paper, be sure to prepare a final version of a replication data set and make it publicly available, such as by submitting it to the ICPSR's Publication Related Archive.

Notes

*My deepest appreciation goes, in addition to my students, to the numerous scholars who have cheerfully, and in some cases repeatedly, responded to my students' queries over many years. Thanks also to the National Institutes of Aging (P01 AG17625-01) and the National Science Foundation (SES-0318275, IIS-9874747) for research support.

1. See King (2003) and http://GKing.Harvard.edu/replication.shtml for more information on the replication and data sharing movement in political science and other fields.

2. The class is Government 2001 at Harvard University. See http://gking.harvard.edu/class.shtml. The course is taken by undergraduates and graduate students from the Government Department and a variety of other departments and schools. An important feature of the class for undergraduates is that they are treated just like graduate students. The graduate students have more wisdom about the literature and what constitute important questions, but the undergraduates often have better mathematical backgrounds or other useful skills. In my experience, the two groups often mesh well together, compete successfully, and can make great coauthor teams, about which more below.

3. If you have a topic that has never before been addressed, it is still best to begin with the closest article to your area. Similarly, major new data collections, while highly desirable generally, are likely to take longer than the time available in a class project and so should be avoided for the purposes of this paper. Even if you ultimately plan a major data collection project, replicating an article at the cutting edge in the literature is usually an excellent place to start. You will learn what is lacking and what might be fixed by your data collection project. You may also be able to gather convincing evidence for potential

https://doi.org/10.1017/S1049096506060252 Published online by Cambridge University Press

funding agencies before you invest a great deal of time.

4. Some partial qualifications: Across all the articles you write after this class and before you come up for tenure, try to coauthor with different people. That way, if there is any question as to your contribution, it will be easy to control appropriately without collinearity. You should probably limit the total number of coauthors on each paper to three or four when possible.

References

Fisher, Bonnie S., Craig T. Cobane, Thomas M. Vander Ven, and Francis T. Cullen. 1998. "How Many Authors Does It Take to Publish an Article? Trends and Patterns in Political Science." *PS: Political Science and Politics* 31(4): 847–856.

Gleditsch, Nils Petter, Claire Metelits, and Havard Strand. 2003. "Posting Your Data: Will You be Scooped or Will You be Famous?" *International Studies Perspectives* 4: 89–97.

Gleditsch, Nils Petter, Patrick James, James Lee Ray, and Bruce Russett. 2003. "Editors' Joint Statement: Minimum Replication Standards for International Relations Journals." *International Studies Perspectives* 4: 105.

Imai, Kosuke, Gary King, and Olivia Lau. 2004. "Zelig: Everyone's Statistical Software." http://gking.harvard.edu/zelig.

King, Gary. 1991. " 'Truth' is Stranger than Prediction, More Questionable Than Causal Inference." *American Journal of Political Science* 35(November): 1047–1053. http://gking.harvard.edu/files/abs/truth-abs.shtml.

———. 1995. "Replication, Replication." *PS: Political Science and Politics* 28(September): 443–499. http://gking.harvard.edu/files/abs/replication-abs.shtml.

———. 2003. "The Future of Replication." *International Studies Perspectives* 4(February): 443–499. http://gking.harvard.edu/files/abs/replvdc-abs.shtml.

King, Gary, James Honaker, Anne Joseph, and Kenneth Scheve. 2001. "Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation." *American Political Science Review* 95(March): 49–69. http://gking.harvard.edu/files/abs/evil-abs.shtml.

King, Gary, and Langche Zeng. 2006. "When Can History Be Our Guide? The Pitfalls of Counterfactual Inference." *International Studies Quarterly*. http://gking.harvard.edu/files/counterf.pdf.

King, Gary, and Michael Laver. 1993. "On Party Platforms, Mandates, and Government Spending." *American Political Science Review* 87(September): 744–750. http://gking.harvard.edu/files/abs/hoff-abs.shtml.

King, Gary, Michael Tomz, and Jason Wittenberg. 2000. "Making the Most of Statistical Analyses: Improving Interpretation and Presentation." *American Journal of Political Science* 44(April): 341–355. http://gking.harvard.edu/files/abs/making-abs.shtml.

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

Tomz, Michael, Jason Wittenberg, and Gary King. 2003. "CLARIFY: Software for Interpreting and Presenting Statistical Results." *Journal of Statistical Software* 8(1). http://gking.harvard.edu/stats.shtml.

https://doi.org/10.1017/S1049096506060252 Published online by Cambridge University Press