UC Berkeley
UC Berkeley Previously Published Works

Title
A framework for open policy analysis

Permalink
https://escholarship.org/uc/item/05r470xk

Journal
Science and Public Policy, 48(2)

ISSN
0302-3427

Authors
de la Guardia, Fernando Hoces
Grant, Sean
Miguel, Edward

Publication Date
2021-06-23

DOI
10.1093/scipol/scaa067

Copyright Information

Peer reviewed

A Framework for Open Policy Analysis

Fernando Hoces de la Guardia[1], Sean Grant[2], Edward Miguel[1,3]

[1] Berkeley Initiative for Transparency in the Social Sciences

[2] Indiana University, Richard M. Fairbanks School of Public Health

[3] University of California, Berkeley, Department of Economics

Abstract word count: 148 words

Paper word count: 6129 words

Abstract

The evidence-based policy movement promotes the use of empirical evidence to inform policy decision-making. While several social science disciplines are undergoing a "credibility revolution" focused on openness and replication, policy analysis has yet to systematically embrace transparency and reproducibility. We argue that policy analysis should adopt the open research practices increasingly espoused in related disciplines to advance the credibility of evidence-based policymaking. We first discuss the importance of evidence-based policy in an era of increasing disagreement about facts, analysis, and expertise. We present a novel framework for "open" policy analysis (OPA) and how to achieve it, focusing on examples of recent policy analyses that have incorporated open research practices such as transparent reporting, open data, and code sharing. We conclude with recommendations on how key stakeholders in evidence-based policy can make OPA the norm and thus safeguard trust in using empirical evidence to inform important public policy decisions.

A Framework for Open Policy Analysis

POLICY ANALYSIS AND THE EVIDENCE-BASED POLICY MOVEMENT

The evidence-based policy movement has gained significant traction over the last two decades. This movement aspires to ensure that decision-making is informed by scientific research and credible evidence (Haskins 2017). Its main focus and major victories thus far have to do with increasing the rigor of the evidence generated: growth in the understanding and production of credible causal evidence (Angrist and Pischke 2010), and a growing emphasis around transparency and reproducibility of research (Miguel et al. 2014). Parallel to these developments in the production of evidence, there has also been a growing demand by some policy makers for such high-quality evidence. Recent efforts in the US Congress – most notably the Foundations for Evidence-Based Policymaking Act of 2018 – demonstrate the growing momentum behind strengthening the federal government's efforts to use data, research, and evaluation for making decisions about government policies and programs (Congress 2018).

However, the process of translating general purpose scientific evidence into advice that contributes to debates about specific policy alternatives (the domain of policy analysis) remains rather obscure. Among the multiple definitions of policy analysis (see, for example, Dunn 2015; Weimer, Vining, and Vining 2017), a common denominator is that of client-oriented empirical analysis specifically commissioned to inform a specific policy debate. In the context of evidence-based policy, policy analysis is a key intermediary between the evidence generated by researchers, designed to contribute to general knowledge via understanding previous policies, and the information required by policy makers to choose among specific policy alternatives. This relationship is depicted in Figure 1. Another key distinction between research and policy analysis is that the former is mainly focused on understanding the effect of previous policies, while the

latter is principally centered around quantifying the effects of future policies (Friedman 2017).

Figure 1 displays a simple diagram of one of the mechanisms through which evidence can connect with policy, building on Nutley et al. (2007). In this ideal context, there is consensus on which research is the best representation of some specific phenomenon (the truth), and which policy analysis is the best representation of gains and losses associated with a specific policy issue (using research as an input). Different policy makers start from a commonly shared set of evidence and make their choices based on a combination of evidence and their different values or beliefs, with these choices revealing each policy makers' values to the public.

Figure 1. Policy-making with high credibility in research and policy analysis
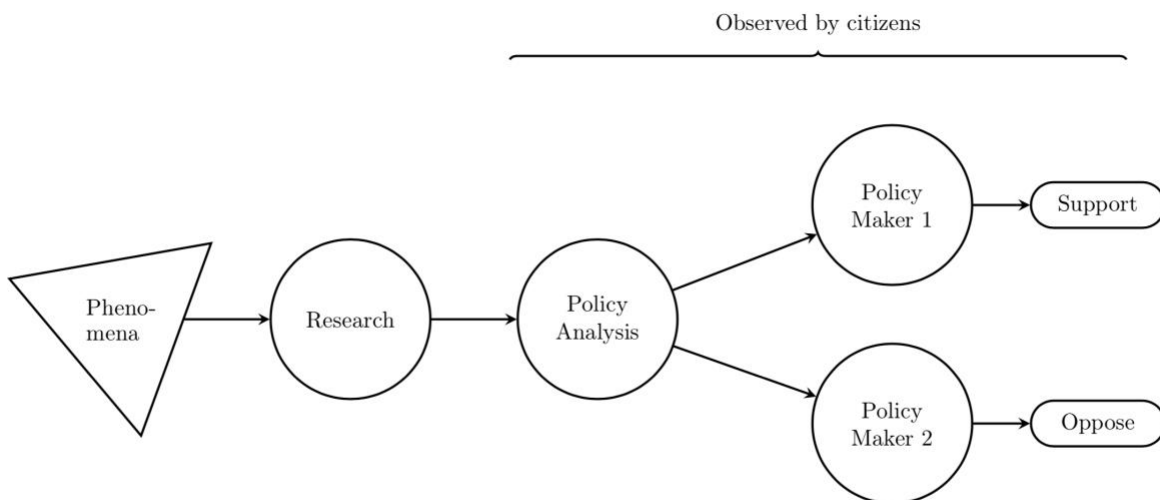

Figure 1: Policy-making with high credibility in research and policy analysis

An illustrative example is the debate around the minimum wage. An ideal evidence-based policy debate around the minimum wage requires credible evidence on the effects of changes in the minimum wage on employment and income. The evidence could describe what would happen to a given population (e.g., working teenagers), given a change in their minimum wage (e.g., $7.25 to $9), on certain outcomes (e.g., employment in a specific industry). Policy makers, however, might want to know about the potential effects of raising the minimum wage for a

different population (e.g., the entire labor force), for a different change in the minimum wage

(e.g., $7.25 to $15), on different outcomes (e.g., employment, income, and distributional effects).

The purpose of policy analysis is to translate evidence from one setting to inform a different one

(for the case of the minimum wage see, for example (Congressional Budget Office 2019)).

Policy analysis aspires to the rigor and neutrality of scientific empirical analysis

(Wildavsky 1979). However, as currently practiced, policy analysis often does not adhere to the

scientific principles of reproducibility and transparency. Until recently, many scientific

disciplines themselves have failed to emphasize reproducibility and openness (Anderson,

Martinson, and De Vries 2007). But as empirical sciences are starting to acknowledge and

address their own reproducibility crises, the need for policy analysis to embrace transparency and

reproducibility has become more evident. In this paper, we argue that opaqueness and

irreproducibility threatens the credibility of policy analysis and therefore evidence-based policy

as a whole. To remedy this concern, we suggest "open science" solutions similar to those

currently being promoted in the empirical sciences.

REPRODUCIBILITY CRISES AND THE OPEN SCIENCE MOVEMENT

The last decade has seen an increase in awareness regarding transparency and

reproducibility across scientific fields. Several problems have been identified, sometimes

grouped under the term "reproducibility crisis". A first concern is researcher degrees of freedom:

the significant latitude and flexibility currently afforded to researchers when planning, running,

analyzing, and reporting empirical research studies (Simmons, Nelson, and Simonsohn 2011;

Wicherts et al. 2016), which may allow them to focus selectively on a subset of results that are

not representative of the evidence as a whole. Second, the file-drawer problem, where positive

results are much more likely to be written up and published than null results (Franco, Malhotra,

and Simonovits 2014; Brodeur et al. 2016). And third, the simple lack of reproducibility: the results of many studies simply cannot be replicated in comparable settings (Open Science Collaboration 2015; Camerer et al. 2016) , or even reproduced computationally using its original data and code (Chang and Li 2015; Gertler, Galiani, and Romero 2018).

In response to these problems, Open Science is defined as "the practice of science in such a way that others can collaborate and contribute, where research data, lab notes and other research processes are freely available, under the terms that enable reuse, redistribution, and reproduction of the research and its underlying data and methods" (Bueno de la Fuente 2017). For our purposes we highlight two key milestones in the larger adoption of open science practices. First, we highlight the articulation of three guiding norms to promote transparency in empirical social sciences: (i) disclosure of details; (ii) registration and pre-analysis plans; and (iii) open access to data and materials (Miguel et al. 2014). Second, we highlight the Transparency and Openness Promotion (TOP) Guidelines (Nosek et al. 2015), which operationalize these norms into policies that journals, funders, and scholarly societies can adopt. The TOP Guidelines contain eight standards, and compliance with each standard is defined on four different levels, with increasing rigor. (For a more detailed discussion of the reproducibility crisis and the contemporary open science movement in the social sciences, refer to the appendix.)

CREDIBILITY CRISIS OF POLICY ANALYSIS

Unlike science, policy analysis does not have a well documented reproducibility crisis. There are, however, related problems that scholars have identified and could be understood under a similar framework of the reproducibility crisis in science. For example, Manski (2013) describes how policy analysis is largely a black box to those outside the analysis team, and policy estimates are nearly always reported with strong yet undisclosed assumptions. Hird (2017)

argues that in practice it is very hard to disentangle policy analysis that is meant to inform from policy advocacy. Finally, the recent rise in the phenomena of fake news (McIntyre 2018) and general distrust in expertise (Nichols 2017) is probably the strongest signal that policy analysis currently also lacks widespread public credibility. More recently, Perl, Howlett, and Ramesh (2018) review the role of policy science in an era of fake news and conclude that "policy scientists and scholars have a responsibility to explain and help society and policy-makers understand policymaking in an era of truthiness and how they can deal with the growth, especially, of willful ignorance and obliviousness."

The lessons from the reproducibility (or credibility) crisis in science, and the response by advocates of open science, can be applied to the policy analysis setting. Policy analysts use the same practices and face the same incentives driving credibility crises in empirical research. When performing a policy analysis, analysts have to define the relevant data, select variables, and choose among multiple methods and definitions. This creates the same problem of a "garden of forking paths", as defined by Gelman and Loken (2013) for empirical research. And, as in academia, incentives are typically mis-aligned to "publish or perish" in policy analysis, leading to undesired consequences that range from not-publishing a result to publishing only those analyses that do not risk the loss of support for public policies with large constituencies behind them.

We argue that the credibility of policy analysis can be greatly improved by adopting open science practices. That is, to facilitate evidence-based policy rather than risk being considered activism or lobbying, the practices, institutions, and scholarly community of policy analysis must aspire to the ideals and ethos of scientific research (Merton 1942). We can now contrast the ideal setting depicted in Figure 1, with an evidence-to-policy model that incorporates low credibility in

both research and policy analysis (see Figure 2). In terms of research inputs, a multiplicity of

studies often exist (R_1, R_2, R_3) that vary in their use of credible research practices (represented by

the solid line) versus questionable research practices (dashed lines). This problem is then

compounded by the fact that policy analysts themselves are subject to a similar set of barriers

and perverse incentives as their research inputs, meaning the "policy analyst" degrees of freedom

must also be incorporated into the model (Simmons, Nelson, and Simonsohn 2011). The result is

a potential multiplicity of reports and conclusions.

Figure 2. Policy-making with low credibility in research and policy analysis
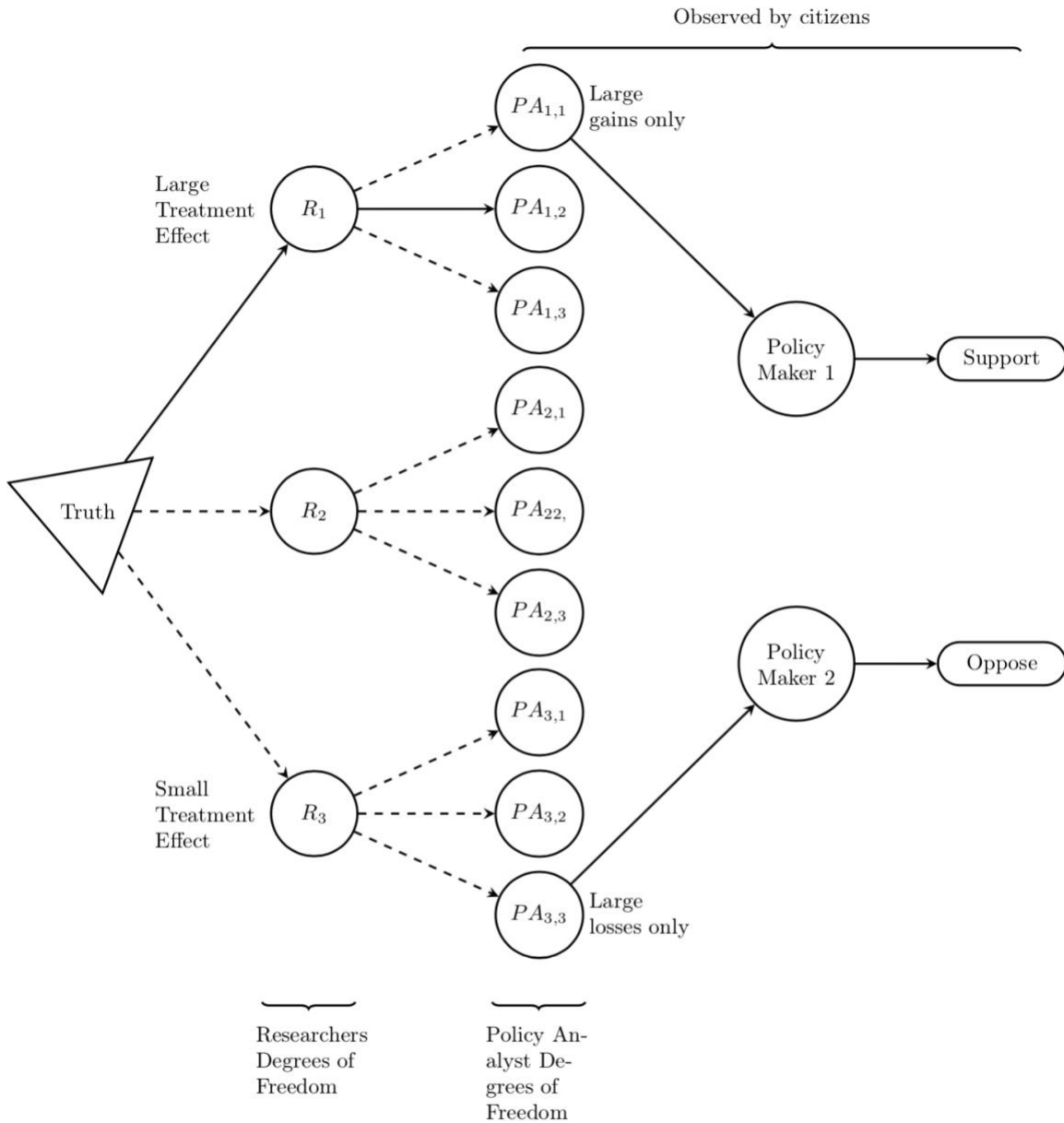

Figure 2: Policy-making with low credibility in research and policy analysis

Using this framework, we can describe at least three problems that emerge from a lack of openness in policy analysis: cherry-picking (weak) evidence, difficulties in automating and improving systematically recurring reports, and difficulties understanding how research informs specific policy analyses.

Cherry-Picking (Weak) Evidence

A plethora of reports and analyses for a given policy question provides policy makers with the opportunity to selectively use (i.e., cherry-pick) the reports and analyses that best fit their pre-selected policy positions, rather than those that are most appropriate or analytically rigorous. Policy makers and other stakeholders often engage in "report wars": i.e., debating the same policy issue based on different empirical policy reports (Wesselink et al. 2013). Even when only one policy report exists, policy makers can selectively highlight those analyses within the report that provide evidence supporting their position and ignore conflicting results. Douglas Elmendorf, former director of the Congressional Budget Office (CBO), provides insight on this issue:

> "When I was director of the CBO, I was very frustrated when we would write a policy report [saying] a certain policy would have these two advantages and these two disadvantages, and the advocates would quote only the part about the advantages, and the opponents would quote only the part about the disadvantages. That encourages the view that there are simple answers. There aren't generally simple answers. There are trade-offs" (Bolotnikova 2016).

These report wars are further compounded when a lack of transparency prevents observers from critically appraising the policy analyses within each report, as policy makers can cherry-pick findings from less neutral and rigorous reports. Moreover, without openness in policy analysis, the credibility of a report largely rests on the reputation of the analysts (Doberstein 2017), in contrast to the dictum from the scientific principle of universalism that the strength of a claim should rest on the quality of the evidence rather than who is making the claim (Merton 1942).

This overall situation threatens the credibility of policy analysis by contributing to

increasing disagreements on analytical interpretations of data, blurring the line between evidence

and opinion, and lowering confidence in respected expert sources of factual information

(McIntyre 2018). For example, the 115th US Congress assigned so little credibility to the

analysis from the Congressional Budget Office (CBO) in their legislative discussion of

healthcare reform that all former CBO directors wrote a letter requesting that Congress give

more weight to CBO's analysis (Crippen et al. 2017). Manski (2013) predicted only a few years

earlier that credibility based solely on reputation would not be sustainable: "I worry that

someday sooner or later the existing social contract to take CBO scores at face value will break

down. Conventional certitudes that lack foundation cannot last indefinitely."

Challenges for Systematically Recurring Reports

A lack of openness in policy analysis can also make it challenging to automate and

improve reports that are intended to systematically recur, leading to an inefficient use of time and

resources. The process of policy analysis can be seen as an algorithmic process. Analysts use

three primary sources for inputs into the analyses: information from previous research (e.g.,

elasticities, behavioral parameters), data to contextualize the specific policy issue (e.g., micro-

data for the specific context where the policy issue is discussed), and guesswork to fill in any

missing pieces required to complete the analysis (e.g., extrapolation parameters, take-up rates,

distributional effects). All of these sources are used to generate inputs that are used in a model

(e.g., micro-simulation, cost benefit analysis), and this model produces the policy estimates to be

used by policy makers.

A large number of policy analyses are recurring reports, where similar analyses are

repeated over a cycle and could potentially benefit from automation. Examples abound across a

wide spectrum of policy domains. For instance, the effects of federal regulations are often

assessed initially and are meant to be updated regularly (Sunstein 2012), a policy analysis is

needed to quantify gains and losses every time there is a discussion about the U.S. minimum

wage (Congressional Budget Office 2007; 2014; 2019), and analyses assessing the

appropriateness of medical treatments should be updated periodically (Zauber et al. 2008; 2016).

Policy reports require a large amount of resources in terms of highly skilled labor. Part of

this work is codified in the final report and can be used in future versions of the analysis.

However, a large amount of tacit knowledge (e.g., how does the spreadsheet/code work? What is

the latest version of the data to be used? How to interpret missing values?) can be lost when the

analyst in charge leaves the institution (or forgets some of the details due to the passage of time).

This is particularly relevant for the guesswork component; whenever the research and data

components do not contain all the relevant information to execute the analysis, some elements of

the model have to be the result of educated guesses made by the analyst. Without a clear record

of the assumptions applied in each iteration of a report, it is almost guaranteed that this

knowledge gets lost across reports. Advances in computational reproducibility can help to

address this issue (Stodden, Leisch, and Peng 2014). Workflow management (Long 2009),

version control (Gentzkow and Shapiro 2014), and dynamic documents (Xie 2017) are among

the main new tools to achieve a one-click-reproducible workflow. Yet, to the best of our

knowledge, these advances are rarely used among policy analysts.

Difficulty Understanding How Research Informs Policy Analysis

Input parameters and specification of models in policy analysis often come from evidence

generated from research. Research can directly inform a policy analysis as a behavioral

parameter obtained from the literature, or indirectly through the modeling choices made in a

policy report. For example, the elasticity of labor demand for a subset of the population is often a

behavioral parameter derived from research when conducting a policy analysis about the gains

and losses from a proposed change in the U.S. minimum wage. A lack of transparency makes it

difficult (if not impossible) to understand precisely what these input values were, whether they

come from credible research, and therefore how well the estimates produced by policy analysis

can be trusted (Nutley, Walter, and Davies 2007; Vivalt 2019). Open policy analysis could help

identify how research informs specific policy analysis models, thereby informing the critical

appraisal of these models as well as value of information exercises for allocating future research

resources to the most important knowledge gaps surrounding specific policies (Snilstveit et al.

2016).

A FRAMEWORK FOR OPEN POLICY ANALYSIS

We propose a conceptual framework of a transparent, reproducible workflow for open

policy analysis. This conceptual framework involves an organized set of principles across the

lifecycle of policy analysis. We suggest three core principles that distinguish open from

traditional policy analysis: open materials, open analysis, and open output (see Figure 3). This

framework seeks to provide guidance for policy analysis concerned with the problems described

above. We created this framework building on the work of several leading groups in research

transparency, namely: the *Lancet* REWARD Campaign (Moher et al. 2016); the Center for Open

Science and their Transparency and Openness Promotion (TOP) Guidelines (Nosek et al. 2015);

the Berkeley Initiative for Transparency in the Social Sciences (Miguel et al. 2014); the Data

Access and Research Transparency (DA-RT) group (Data Access and Research Transparency

group 2015); the Meta-Research Innovation Center at Stanford (J. P. A. Ioannidis et al. 2015);

and the Laura and John Arnold Foundation (Preston 2011).

Figure 3. A Conceptual Framework for Open Policy Analysis
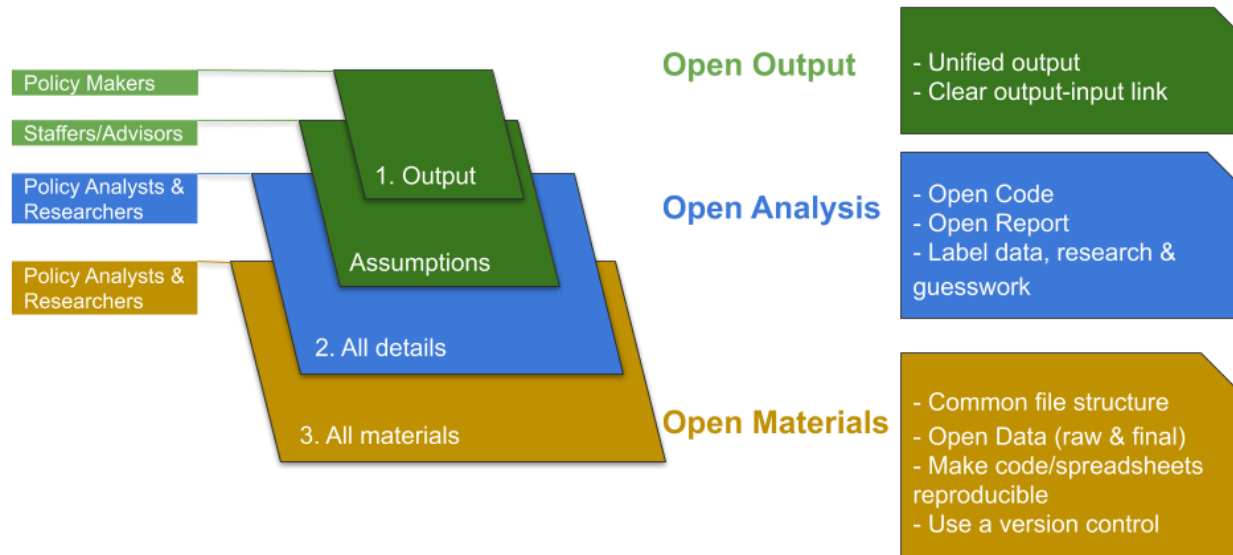
We propose that, in contrast to a traditional policy analysis contained in a printed report, an OPA should be presented in four layers: the top layer is the final output to be presented to policy makers. Policy makers have to weigh the different benefits and costs to make their decisions. However, they should not in general have leeway to choose what is the best representation of those benefits and costs. Even in settings with considerable uncertainty, it is the responsibility of the policy analysts to identify what is the best representation of the facts in a report. In a second layer, the staffers and advisors of policy makers should be able to observe what the main assumptions of the analysis are, and how sensitive the results are to changes in those assumptions.

The documentation of the report corresponds to a third layer. The intended audience of this third layer would be other policy analysts as well as researchers interested in gaining an in-depth understanding of the estimates produced in the first two layers. Finally, the bottom foundational layer should contain all the elements required to reproduce, with the least possible effort, all the results presented in the other layers. The intended audience for this layer are also policy analysts (including the original authors of future iterations of the report) and researchers

interested in reproducing the analysis from beginning to end.

With this layered structure, we can now elaborate on the three organizing principles for OPA:

1. *Open Output*: The analysis should clearly (pre-)specify the output that will inform policy makers, identify the preferred set of estimates, and properly communicate underlying uncertainties, as well as how key results vary with underlying assumptions.

2. *Open Analysis*: All elements of the analysis should be easily accessible and readable for critical appraisal and improvement. This includes disclosing all methodological procedures and underlying assumptions behind the report.

3. *Open Materials*: Raw data, code, and supporting documents should be made publicly available, to the extent legally possible, to allow a policy report to be reproduced in its entirety with the least possible effort. This principle is akin to the concept of computational reproducibility.

To operationalize each principle, we suggest the specific dimensions below to assess the openness of any given policy analysis.

Dimensions for an Open Output

Minimal and clear output display (ideally pre-committed). Policy makers often do not have the time to read full reports or lengthy executive summaries, and may instead prefer concise and non-technical briefs. Moreover, when presented with several scenarios, policy makers are more likely to cherry-pick their preferred estimate. For these reasons, we propose that policy analysts should present one final output as the best representation of the facts. If time allows, they could pre-specify the nature of table or visualization that contains the different gains and losses associated with a particular policy. Then, the format of such output, together with the

methodology of the report, could be vetted before the final output is released. This ensures that the community of academics and experts judge the merits of the report independently of the results, and reduces the chance that features other than the rigor of the analyses (e.g., the implications of the output) influence the methodological credibility of the report. If time constraints do not allow for such a pre-committed format, the report should refer to output used in previous versions of a similar analysis and justify deviations. In this case, the new produced output will become the default pre-specified format for the next iteration of a similar analysis.

Clear assumptions-output link. For an agreed format, policy analysts can also report and display the results for different values of the key assumptions behind the analysis, within a range of scientifically credible values. This can be done by adding pages to the report with the respective sensitivity analyses (with all outputs reported in a consistent format), or by displaying an interactive output (e.g., using R Shiny) that changes when the assumptions are modified. The latter makes it easier to identify the preferred set of assumptions for each reader and, even if different policy makers prefer distinct assumptions, this format still reduces the discrepancies from the "report wars" discussed in the previous section, to differences in the values of specific and ideally testable parameters.

Dimensions for an Open Analysis

All the elements of the analysis should be easily accessible (and readable) for critical appraisal and improvement. Here, some of the specific practices from open science that apply to open policy analysis are:

Open code: all the files used in the analysis (including data cleaning) should be available in a trusted public repository, to the extent legally possible. To increase accessibility, the practice of code readability described above should also be considered. The recent American Economic

Association guidelines for data and code sharing provide a useful recent template (American Economic Association, 2019).

Open Report: a detailed methodological description is essential to critically appraise and improve the methodology. The report should describe the analysis in increasing layers of detail: from a first layer that describes the key assumptions and results of different subsections of the report, to a last layer that combines a detailed description of each step and how it is implemented in the code. This layered approach can be implemented as a Dynamic Document (Xie 2017)(Xie 2017), where code, narrative, and output are in the same file, and the reader can expand each section to see every piece of the analysis. Even though the foundational concepts behind dynamic documents were established decades ago (Knuth 1984), the implementation of dynamic documents is still in its infancy, and some languages/software packages are more developed than others. To our knowledge, the implementations for R (RMarkdown) and Python (Jupyter notebooks) are currently the ones best suited for reproducibility, although even some commercial proprietary packages, such as STATA, are beginning to incorporate these concepts.

Label and document each input sources (data, research, guesswork): in order to improve the reproducibility of previous reports, we recommend that a labeling strategy should be in place to trace all the inputs of each policy analysis. Similar to empirical research, this labeling strategy involves identifying sources of data and previous research. Specifically, to policy analysis, this labeling strategy also requires keeping track of the parameters that required an educated guess to complete the analysis. This record should include detailed information on each source. Inputs from research should record the paper, page, and specific location (i.e., table X, row Y, col Z) from which the piece of information was obtained. For data, links to specific data sets should be provided. For guesswork, parameters should be recorded similarly as with research (analyst X,

date Y).

Dimensions for Open Materials

As part of open materials (or computational reproducibility), policy analysts should follow best practices for a fully reproducible workflow. Some of the practices used in empirical research that apply directly to the policy analysis setting are:

Open data: all the data used in the analysis should be posted in a trusted repository when possible. This includes both the raw data and the final analytic files. Whenever some of the data cannot be published due to privacy and ethical issues, a clear protocol should describe all the procedures required to obtain access to a subset of these data (e.g., instructions on obtaining restricted access to the dataset). For access to administrative government data, we highlight three of the recommendations from the Commission on Evidence-Based Policymaking (Abraham et al. 2018) that are related to the subsequently enacted Foundations for Evidence-Based Policymaking Act (Congress 2018). First, the government should provide proper infrastructure to access sensitive information for research purposes. Second, state-collected quarterly earnings and other local data on federal programs should be made available in as findable, accessible, interoperable, and reusable a manner as possible (Wilkinson et al. 2016). Third, a uniform process should be in place for external researchers to apply and qualify for secure access to confidential government data.

Make code/spreadsheets readable: an old idea that has gained traction in reproducible research is that of making the code readable not only by machines but also by humans (Knuth 1984; Stodden et al. 2016). We envision that this idea can also be applied to analysis using spreadsheets, by providing extensive comments and adopting some type of standard operating procedures in how the analysis is carried out (Lin and Green 2016). Applied to code, this is the

principle of "literate programming", and it amounts to describing what is happening in each line (or few lines) of code in some detail. Applying this principle to analysis using spreadsheets is less well defined, but we envision some type of syntax guideline that would help the analyst comment on the meaning of the different components of a spreadsheet analysis. Whether in code or in spreadsheets, the documentation should provide an explanation to key elements of the analysis, such as the rationale behind the modeling choices and a description of all analytical assumptions.

Use a version control strategy: to keep track of how files change over time, a systematic and homogenous strategy should be in place (as opposed to some idiosyncratic file renaming strategy that varies across analysts). Plausible strategies range from a minimum organized renaming to the (much recommended) use of distributed version control (Ram 2013). Organized renaming involves setting a file naming convention (e.g. YYYYMMDD_filename_initials) and saving on a new file periodically. Using version control software, like Git or SubVersion, involves using only one file per document (no more dates or initials in filename) and having snapshots of the entire workflow at much higher frequency than the renaming strategy.

Follow a common template for file structure: one of the most common recommendations, and one with a low adoption cost, is that all analysts should follow some homogenous and predetermined file structure. In practice, this means that multiple analysts within a group store their workflow in an identical way and with all the files under a master folder (for example see Gentzkow and Shapiro 2014).

AN APPLICATION TO A POLICY ANALYSIS FOR A WEALTH TAX IN THE US

To illustrate this framework, we present a case study for OPA based on an original policy analysis by Saez and Zucman (2019) to assess the fiscal effects of a wealth tax in the US. The

OPA, in this case, consisted of three separate components: an interactive visualization, a dynamic document, and a public repository that contains all the materials to reproduce the analysis with the least possible effort.

The interactive visualization was built using R statistical software (and its library Shiny). It displays the main policy estimate--the total revenue from a given wealth tax schedule--in an interactive fashion. This means that the user of the tool can modify both the key assumptions of the analysis (e.g., the tax avoidance rate) and its design features (income brackets and tax rates), and immediately see how the total revenue generated changes. This corresponds to the principle of *open output.* Figure 4a shows a screenshot of the tool (available at

http://wealthtaxsimulator.org/simulator_app/).

Figure 4: Open Output and Open Analysis in Wealth Tax Policy Analysis
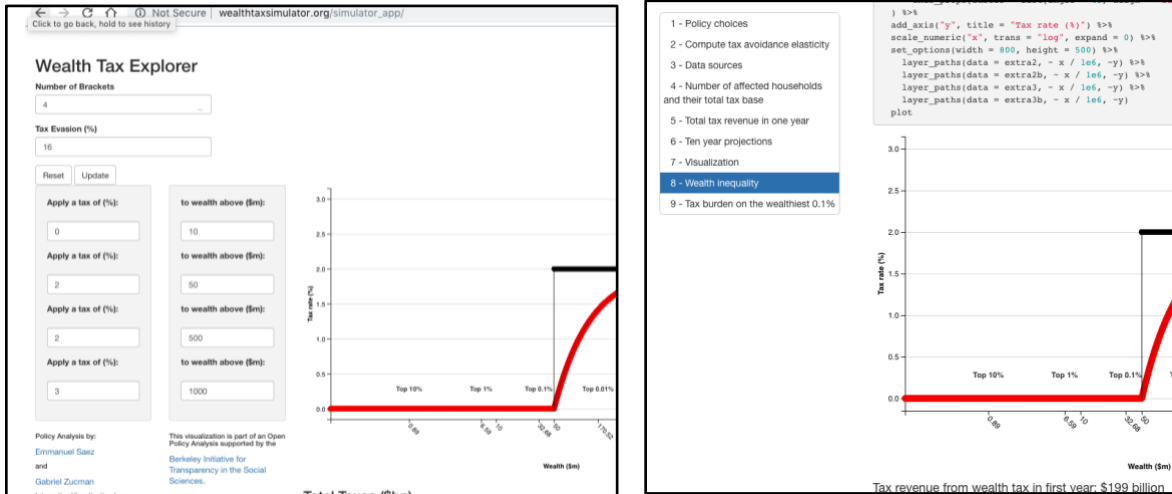
Panel a: Open Output   Panel b: Open Analysis

The second component is a dynamic document that explains in complete detail the analysis and code underlying the policy estimates. This report (like all dynamic documents) combines the narrative of the report with the statistical code required for its analysis, and with its corresponding output, in a single document (in this case, as a website format where code can be unfolded, but output can also be produced as a pdf or word document, without code unfolding). All the numbers in this report are generated programmatically (as opposed to typed manually), hence allowing for the interested reader to trace exactly the origin of all the elements in the analysis. This corresponds to the principle of *open analysis*. Figure 4b shows a screenshot of the report (available at http://wealthtaxsimulator.org/analysis/).

In the final component of this OPA, researchers and policy analysts can find all the materials needed to reproduce the two components above with a few clicks on their own computer. The materials are hosted using version control software (Git), such that any user can create their own copy and build a complete history of changes to track the differences between

the two reports. This corresponds to the principle of *open materials* (available at

https://github.com/BITSS/opa-wealthtax).

A Contrast Between The Wealth Tax OPA and Traditional Policy Analysis

As the policy proposal supported by this analysis has been heavily debated in the US,

competing analyses emerged, providing a unique opportunity to contrast the traditional style of

policy analysis with OPA as proposed here. Specifically, the high-profile analysis in Sarin and

Summers (2019) provides a much smaller policy estimate of revenue generated from a wealth

tax: $25-75 billion, as opposed to the $187 billion implied by the original policy analysis in Saez

and Zucman. These discrepancies could lead an interested policy analyst or researcher to ask

about key inputs and analytical choices that are not discussed, and how the estimates differ by

revising model parameters. For example minor modifications in the parameters used by Sarin

and Summers would have produced estimates much closer to the original analysis[1]. In addition,

per the concerns motivating our OPA framework and consistent with the core scientific ethos

(Merton 1942), analysts should avoid refuting differing analyses through rhetorical exercises

(rather than methodological critiques), appeals to the authority of the analyst (rather than the

quality of the underlying work), and selectively focusing on analyses that support one's position

and ignoring equally methodologically valid analyses that run counter to it.

---

[1]Using age 60 instead of 70 to compute mortality rates and an effective tax rate of 13% instead of 13.5%
US$ 167 billion in 2017 (as oppose to their original 75 billion) or 186 billion in 2019 dollars when adjusting
the 10 million dollars collected by the state tax, and using the same growth rate as in the original study
$(10*(1+0.055)^2/(0.0092 * 0.13) * 0.02)$.

EXPECTED BENEFITS, LIKELY BARRIERS, AND NEXT STEPS

Expected Benefits

The OPA framework proposed here is meant to address the threats to the credibility of policy analysis. *Open output* would drastically reduce the ability of policy makers to cherry-pick evidence within a report by condensating all the relevant costs and benefits in one clear output for all policy makers. *Open analysis* and *open materials* would allow other policy analysts (including the original authors themselves in the future) to reproduce the analysis, and improve upon it, ideally with minimal effort. And compliance with all three principles would allow researchers to easily track how their research is used in a policy analysis and, more importantly, identify the prospective effect that a new study could have on the current estimates. For example, researchers could define their minimal detectable effect sizes from a power calculation up front, with the final effect on the policy estimates as a target. All of these potential improvements would buttress high quality policy analyses that appear to be under increasing attack due to more general public skepticism regarding the role of expertise (Nichols 2017).

Likely Barriers

While we believe open policy analysis will have sizable benefits, the barriers to its adoption are similar to the ones recently faced by open science in other research areas (Pfenninger et al. 2017). Policy makers may not want analyses to be open when arguing for highly scrutinized and partisan issues if openness could threaten their predetermined policy positions or decisions. In addition, analysts may have other reasons to prefer keeping policy analyses largely "closed". When performing under tight deadlines for a specific client, the human errors likely to be made are discoverable only if open science practices are used, and disclosure of these errors could lead to embarrassment for the analyst. Moreover, much policy

analysis is contracted out to private third parties, where revealing methods may compromise their business interests relative to competitors. Making a policy report "open" will also add an additional layer of work up front (even if it speeds up future work). For example, analysts who do not already document the steps to their analysis in a detailed and reproducible manner will need time and training to develop these new skills, and further time to fully incorporate them into their workflows. Institutional decisions on project timelines may also need to be modified to accommodate new open policy analysis activities. While the above concerns may be addressable in the longer-term, another standing concern is that some data may not be shareable in their analyzed form due to the sensitivity of information -- for instance, individual tax or medical data -- necessitating additional time and resources for adequate de-identification, or alternative forms of data access, if open data is expected.

All these likely barriers can be framed under a set of "anti-Mertonian" research norms, such as those described by (Ziman 1994). Many researchers and policy analysts have strong incentives to make their work proprietary, local (as opposed to universal), authority-based, commissioned, and expert (as opposed to disinterested, and based on organized skepticism). Some of these anti-Mertonian norms are inherent to policy research, namely "commissioned research", since policy analysis is typically commissioned to inform a particular stakeholder. Moreover, the absence of external peer review in most policy analysis risks heightening the role of authority and the status of the authors in assessing the output of policy analyses, which runs against the Mertonian norm of universality. As with reforms in the empirical social sciences, explicitly identifying each of these likely barriers as specific anti-Mertonian norms can facilitate how they can be addressed and overcome to enable OPA. In other words, as with open science among academic scholars, reforming the institutions and incentives that govern policy analysts'

behavior is the key to overcoming current challenges (National Academies of Sciences 2018).

Next Steps

Future research is needed to investigate the implementation and impact of OPA as more cases occur in practice. For example, several organizations are already developing analyses that are aligned with the framework proposed here. The Open Source Policy Center has an open source tax reform simulator (OSPC 2018). Their interactive tool allows users to explore variations in assumptions of the model and policies of interest, with all their code available online. In another notable example, the charity evaluator GiveWell has open spreadsheets for their cost benefit analyses, allowing the public to examine what goes into their analysis (GiveWell 2017). Other examples include online interactive tools that display multiple policy scenarios and emphasize uncertainty in estimates (L. R. Gerber et al. 2018). Empirical research on these and other cases can examine how OPA affects the quality of proposed policies, whether any actors challenge proposed policies using the open materials and analyses that traditionally have been unavailable, and whether implementers or enactors of policies use OPA to fine tune policy proposals.

Following the example of the open science movement, OPA also could benefit from defining guidelines for policy analysts and funders, developing more case studies that showcase how to do OPA in practice, and building a community of practice. A proposed set of guidelines, paired with a demonstration of tools like dynamic documents and open source code for the case of the minimum wage in the US, can be found in (Hoces de la Guardia 2017). For more detail, policy analysts and other stakeholders can consult the resources page of the Berkeley Initiative for Transparency in the Social Sciences (BITSS) for up-to-date information on practicing open science (www.bitss.org/opa/).

CONCLUSION

Open science has, in less than a decade, moved some of the above suggestions from abstract ideals to widespread or even mandatory practices in certain branches of empirical social science research, such as development economics (Christensen et al. 2019). In this paper, we have argued that policy analysis can address its credibility crisis by adopting solutions from the open science movement. The emergence of a suite of new transparency tools and approaches means that now is a time when the policy analysis community can realistically embrace open policy analysis. Key principles of open policy analysis involve having transparent and reproducible outputs, analyses, and materials. Future work is needed to investigate the effectiveness of OPA in improving the quality of the resulting proposed and adopted policies. We hope that the framework laid out in this article identifies the key issues and provides a direction to move OPA forward.

REFERENCES

Abraham, KG, R Haskins, Sherry Glied, RM Groves, Robert Hahn, Hilary Hoynes, and KR Wallin. 2018. "The Promise of Evidence-Based Policymaking: Report of the Commission on Evidence-Based Policymaking." *Washington, DC: Commission on Evidence-Based Policymaking*.

Anderson, Melissa S., Brian C. Martinson, and Raymond De Vries. 2007. "Normative Dissonance in Science: Results from a National Survey of u.s. Scientists." *Journal of Empirical Research on Human Research Ethics: JERHRE* 2 (4): 3–14. https://doi.org/10.1525/jer.2007.2.4.3.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2010. "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics." *Journal of Economic Perspectives* 24 (2): 3–30. https://doi.org/10.1257/jep.24.2.3.

Baker, Monya. 2016. "1,500 Scientists Lift the Lid on Reproducibility." *Nature News* 533 (7604): 452. https://doi.org/10.1038/533452a.

Bolotnikova, Marina N. 2016. "A Moral Conscience for Economics." Harvard Magazine. August 9, 2016. https://harvardmagazine.com/2016/09/a-moral-conscience-for-economics.

Brodeur, Abel, Mathias Lé, Marc Sangnier, and Yanos Zylberberg. 2016. "Star Wars: The Empirics Strike Back." *American Economic Journal: Applied Economics* 8 (1): 1–32. https://doi.org/10.1257/app.20150044.

Bueno de la Fuente, Gema. 2017. "What Is Open Science? Introduction,." FOSTER FACILITATE OPEN SCIENCE TRAINING FOR EUROPEAN RESEARCH. 2017. https://www.fosteropenscience.eu/node/1420.

Camerer, Colin F., Anna Dreber, Eskil Forsell, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, et al. 2016. "Evaluating Replicability of Laboratory Experiments in Economics." *Science* 351 (6280): 1433–36. https://doi.org/10.1126/science.aaf0918.

Chang, Andrew C., and Phillip Li. 2015. "Is Economics Research Replicable? Sixty Published Papers from Thirteen Journals Say 'Usually Not.'" SSRN Scholarly Paper ID 2669564. Rochester, NY: Social Science Research Network. https://papers.ssrn.com/abstract=2669564.

Christensen, Garret, Jeremy Freese, and Edward Miguel. 2019. *Transparent and Reproducible Social Science Research: How to Do Open Science*. University of California Press.

Christensen, Garret, Zenan Wang, Elizabeth Levy Paluck, Nicholas Swanson, David J. Birke, Edward Miguel, and Rebecca Littman. 2019. "Open Science Practices Are on the Rise: The State of Social Science (3S) Survey." Preprint. MetaArXiv. https://doi.org/10.31222/osf.io/5rksu.

Congress, US. 2018. "Foundations for Evidence-Based Policy Making Act of 2018." In . Vol. 4174.

Congressional Budget Office. 2007. "Response to a Request by Senator Grassley About the Effects of Increasing the Federal Minimum Wage Versus Expanding the Earned Income Tax Credit | Congressional Budget Office." 2007. https://www.cbo.gov/publication/18281.

———. 2014. "The Effects of a Minimum-Wage Increase on Employment and Family Income | Congressional Budget Office." 2014. https://www.cbo.gov/publication/44995.

———. 2019. "The Effects on Employment and Family Income of Increasing the Federal Minimum Wage | Congressional Budget Office." 2019. https://www.cbo.gov/publication/55410.

Crippen, DL, DW Elmendorf, D Holtz-Eakin, JE O'Neill, PR Orszag, and RG Penner. 2017. "Letter from Former CBO Directors on the Importance of CBO's Role in the Legislative Process. Retrieved 2018 February 19."

Data Access and Research Transparency group. 2015. "Data Access and Research

Transparency (DA-RT): A Joint Statement by Political Science Journal Editors." *Political Science Research and Methods* 3 (3): 421–421. https://doi.org/10.1017/psrm.2015.44.

Doberstein, Carey. 2017. "Whom Do Bureaucrats Believe? A Randomized Controlled Experiment Testing Perceptions of Credibility of Policy Research." *Policy Studies Journal* 45 (2): 384–405. https://doi.org/10.1111/psj.12166.

Dunn, William N. 2015. *Public Policy Analysis*. Routledge.

Franco, Annie, Neil Malhotra, and Gabor Simonovits. 2014. "Publication Bias in the Social Sciences: Unlocking the File Drawer." *Science* 345 (6203): 1502–5.

Friedman, Lee S. 2017. "Public Policy Making and Public Policy Analysis." In *Does Policy Analysis Matter?*, edited by Lee S. Friedman, 1st ed., 1–43. Exploring Its Effectiveness in Theory and Practice. University of California Press. http://www.jstor.org/stable/10.1525/j.ctt1kc6jxn.6.

Gelman, Andrew, and Eric Loken. 2013. "The Garden of Forking Paths: Why Multiple Comparisons Can Be a Problem, Even When There Is No 'Fishing Expedition' or 'p-Hacking' and the Research Hypothesis Was Posited Ahead of Time." *Department of Statistics, Columbia University*.

Gentzkow, Matthew, and Jesse M Shapiro. 2014. "Code and Data for the Social Sciences: A Practitioner's Guide." *Chicago, IL: University of Chicago*.

Gerber, Alan, and Neil Malhotra. 2008. "Do Statistical Reporting Standards Affect What Is Published? Publication Bias in Two Leading Political Science Journals." *Quarterly Journal of Political Science* 3 (3): 313–26.

Gerber, Leah R., Michael C. Runge, Richard F. Maloney, Gwenllian D. Iacona, C. Ashton Drew, Stephanie Avery-Gomm, James Brazill-Boast, et al. 2018. "Endangered Species Recovery: A Resource Allocation Problem." *Science* 362 (6412): 284–86. https://doi.org/10.1126/science.aat8434.

Gertler, Paul, Sebastian Galiani, and Mauricio Romero. 2018. "How to Make Replication the Norm." *Nature* 554 (7693): 417–19. https://doi.org/10.1038/d41586-018-02108-9.

GiveWell. 2017. "Cost-Effectiveness." GiveWell. 2017. https://www.givewell.org/how-we-work/our-criteria/cost-effectiveness.

Haskins, Ron. 2017. "Presidential Address: Making Federal Social Programs Work." *Journal of Policy Analysis and Management* 36 (2): 276–302.

Hird, J. 2017. "How Effective Is Policy Analysis." *Does Policy Analysis Matter*, 44–84.

Hoces de la Guardia, Fernando. 2017. "How Transparency and Reproducibility Can Increase Credibility in Policy Analysis: A Case Study of the Minimum Wage Policy Estimate." RAND.

Ioannidis, John P. A., Daniele Fanelli, Debbie Drake Dunne, and Steven N. Goodman. 2015. "Meta-Research: Evaluation and Improvement of Research Methods and Practices." *PLOS Biology* 13 (10): e1002264. https://doi.org/10.1371/journal.pbio.1002264.

Ioannidis, John PA. 2005. "Why Most Published Research Findings Are False." *PLos Med* 2 (8): e124.

———. 2018. "Meta-Research: Why Research on Research Matters." *PLoS Biology* 16 (3): e2005468.

John, Leslie K, George Loewenstein, and Drazen Prelec. 2012. "Measuring the Prevalence of Questionable Research Practices with Incentives for Truth Telling." *Psychological Science* 23 (5): 524–32.

Kitzes, Justin, Daniel Turek, and Fatma Deniz. 2017. *The Practice of Reproducible Research: Case Studies and Lessons from the Data-Intensive Sciences*. Univ of California Press.

Knuth, Donald Ervin. 1984. "Literate Programming." *The Computer Journal* 27 (2): 97–111.

LeBel, Etienne P, Wolf Vanpaemel, RJ McCarthy, BD Earp, and Malte Elson. 2017. "A Unified Framework to Quantify the Trustworthiness of Empirical Research." *Retrieved from Psyarxiv. Com/Uwmr8*.

Lin, Winston, and Donald P Green. 2016. "Standard Operating Procedures: A Safety Net for Pre-Analysis Plans." *PS: Political Science & Politics* 49 (3): 495–500.

Long, J Scott. 2009. *The Workflow of Data Analysis Using Stata*. Stata Press College Station, TX.

Manski, Charles F. 2013. *Public Policy in an Uncertain World: Analysis and Decisions*. Harvard University Press.

McIntyre, Lee. 2018. *Post-Truth*. MIt Press.

Merton, Robert K. 1942. "A Note on Science and Democracy." *J. Legal & Pol. Soc.* 1: 115.

Miguel, Edward, Colin Camerer, Katherine Casey, Joshua Cohen, Kevin M Esterling, Alan Gerber, Rachel Glennerster, Don P Green, Macartan Humphreys, and Guido Imbens. 2014. "Promoting Transparency in Social Science Research." *Science* 343 (6166): 30–31.

Moher, David, Paul Glasziou, Iain Chalmers, Mona Nasser, Patrick MM Bossuyt, Daniël A Korevaar, Ian D Graham, Philippe Ravaud, and Isabelle Boutron. 2016. "Increasing Value and Reducing Waste in Biomedical Research: Who's Listening?" *The Lancet* 387 (10027): 1573–86.

National Academies of Sciences, Engineering, and Medicine. 2018. *Open Science by Design: Realizing a Vision for 21st Century Research*. Washington, DC: The National Academies Press. https://doi.org/10.17226/25116.

Nichols, Tom. 2017. *The Death of Expertise: The Campaign against Established Knowledge and Why It Matters*. Oxford University Press.

Nosek, B. A., G. Alter, G. C. Banks, D. Borsboom, S. D. Bowman, S. J. Breckler, S. Buck, et al. 2015. "Promoting an Open Research Culture." *Science* 348 (6242): 1422. https://doi.org/10.1126/science.aab2374.

Nutley, Sandra M., Isabel Walter, and Huw T.O. Davies. 2007. *Using Evidence: How Research Can Inform Public Services*. 1st ed. Bristol University Press. https://doi.org/10.2307/j.ctt9qgwt1.

Open Science Collaboration. 2015. "Estimating the Reproducibility of Psychological Science." *Science* 349 (6251): aac4716. https://doi.org/10.1126/science.aac4716.

OSPC. 2018. "Open Source Policy Center." 2018. https://www.ospc.org/.

Perl, Anthony, Michael Howlett, and M. Ramesh. 2018. "Policy-Making and Truthiness: Can Existing Policy Models Cope with Politicized Evidence and Willful Ignorance in a 'Post-Fact' World?" *Policy Sciences* 51 (4): 581–600. https://doi.org/10.1007/s11077-018-9334-4.

Pfenninger, Stefan, Joseph DeCarolis, Lion Hirth, Sylvain Quoilin, and Iain Staffell. 2017. "The Importance of Open Data and Software: Is Energy Research Lagging Behind?" *Energy Policy* 101 (February): 211–15. https://doi.org/10.1016/j.enpol.2016.11.046.

Preston, Caroline. 2011. "A Thirtysomething Billionaire Couple's Bold Philanthropy." The Chronicle of Philanthropy. October 16, 2011. https://www.philanthropy.com/article/A-Thirtysomething-Billionaire/157613/.

Ram, Karthik. 2013. "Git Can Facilitate Greater Reproducibility and Increased Transparency in Science." *Source Code for Biology and Medicine* 8 (1): 7. https://doi.org/10.1186/1751-0473-8-7.

Saez, Emmanuel, and Gabriel Zucman. 2019. "To Senator Elizabeth Warren," 2019.

Sarin, Natasha, and Lawrence H. Summers. 2019. "Opinion | A 'Wealth Tax' Presents a Revenue Estimation Puzzle." Washington Post. 2019. https://www.washingtonpost.com/opinions/2019/04/04/wealth-tax-presents-revenue-estimation-puzzle/.

Simmons, Joseph P., Leif D. Nelson, and Uri Simonsohn. 2011. "False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant." *Psychological Science* 22 (11): 1359–66.

https://doi.org/10.1177/0956797611417632.

Snilstveit, Birte, Martina Vojtkova, Ami Bhavsar, Jennifer Stevenson, and Marie Gaarder. 2016. "Evidence & Gap Maps: A Tool for Promoting Evidence Informed Policy and Strategic Research Agendas." *Journal of Clinical Epidemiology* 79: 120–29. https://doi.org/10.1016/j.jclinepi.2016.05.015.

Stodden, Victoria, Friedrich Leisch, and Roger D Peng. 2014. *Implementing Reproducible Research*. CRC Press.

Stodden, Victoria, Marcia McNutt, David H. Bailey, Ewa Deelman, Yolanda Gil, Brooks Hanson, Michael A. Heroux, John P. A. Ioannidis, and Michela Taufer. 2016. "Enhancing Reproducibility for Computational Methods." *Science* 354 (6317): 1240–41. https://doi.org/10.1126/science.aah6168.

Sunstein, Cass R. 2012. "The Office of Information and Regulatory Affairs: Myths and Realities." SSRN Scholarly Paper ID 2192639. Rochester, NY: Social Science Research Network. https://papers.ssrn.com/abstract=2192639.

Vivalt, Eva. 2019. "How Much Can We Generalize from Impact Evaluations?," 72.

Weimer, David L., Aidan R. Vining, and Aidan R. Vining. 2017. *Policy Analysis : Concepts and Practice*. Routledge. https://doi.org/10.4324/9781315442129.

Wesselink, Anna, Karen S. Buchanan, Yola Georgiadou, and Esther Turnhout. 2013. "Technical Knowledge, Discursive Spaces and Politics at the Science–Policy Interface." *Environmental Science & Policy*, SI: Environmental and Developmental Discourses: Technical knowledge, discursive spaces and politics, 30 (June): 1–9. https://doi.org/10.1016/j.envsci.2012.12.008.

Wicherts, Jelte M., Coosje L. S. Veldkamp, Hilde E. M. Augusteijn, Marjan Bakker, Robbie C. M. van Aert, and Marcel A. L. M. van Assen. 2016. "Degrees of Freedom in Planning, Running, Analyzing, and Reporting Psychological Studies: A Checklist to Avoid p-Hacking." *Frontiers in Psychology* 7 (November). https://doi.org/10.3389/fpsyg.2016.01832.

Wildavsky, Aaron. 1979. *Speaking Truth to Power: The Art and Craft of Policy Analysis*. Speaking Truth to Power: The Art and Craft of Policy Analysis. Piscataway, NJ, US: Transaction Publishers.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3 (1): 1–9. https://doi.org/10.1038/sdata.2016.18.

Xie, Yihui. 2017. *Dynamic Documents with R and Knitr*. Chapman and Hall/CRC.

Zauber, Ann G., Amy B. Knudsen, Carolyn M. Rutter, Steffie K. Naber, V. Paul Doria-Rose, Chester Pabiniak, Colden Johanson, Sara E. Fischer, Iris Lansdorp-Vogelaar, and Karen M. Kuntz. 2016. "Estimation of Benefits, Burden, and Harms of Colorectal Cancer Screening Strategies." *JAMA* 315 (23): 2595–2609. https://doi.org/10.1001/jama.2016.6828.

Zauber, Ann G., Iris Lansdorp-Vogelaar, Amy B. Knudsen, Janneke Wilschut, Marjolein van Ballegooijen, and Karen M. Kuntz. 2008. "Evaluating Test Strategies for Colorectal Cancer Screening: A Decision Analysis for the U.S. Preventive Services Task Force." *Annals of Internal Medicine* 149 (9): 659–69. https://doi.org/10.7326/0003-4819-149-9-200811040-00244.

Ziman, John M. 1994. *Prometheus Bound*. Cambridge University Press.

APPENDIX 1: BACKGROUND ON THE REPRODUCIBILITY CRISIS AND THE OPEN SCIENCE MOVEMENT

REPRODUCIBILITY CRISIS IN EMPIRICAL RESEARCH

Several concerns about the credibility of empirical research have been identified in scientific disciplines using research methods and practices that are commonplace in policy analysis. That is, recent evidence suggests that several areas of empirical research exhibit low compliance with core scientific principles (Anderson, Martinson, and De Vries 2007). This low compliance has serious negative consequences for the credibility of the scientific output underpinning "evidence-based" movements across scientific fields. Some of this discussion draws on Hoces de la Guardia (2017).

Questionable Research Practices and Researcher Degrees of Freedom

In contrast to scientific misconduct or outright fraud, questionable research practices (QRPs) denote the much more prevalent grey areas of scientific practice that are currently acceptable yet nevertheless can dramatically increase the likelihood of finding evidence in support of a (desired) hypothesis (John, Loewenstein, and Prelec 2012). QRPs are facilitated by "*Researcher Degrees of Freedom*": the significant latitude and flexibility currently afforded to researchers when planning, running, analyzing, and reporting empirical research studies (Simmons, Nelson, and Simonsohn 2011; Wicherts et al. 2016). These mechanisms make it unacceptably easy for researchers to (consciously or unconsciously) report a desired rather than accurate result, and therefore for entire bodies of research to have far more false-positive findings than previously thought.

File Drawer Problem and Publication Bias

In the last decade, "meta-research", or research on research, has become an active field

across multiple disciplines (Ioannidis 2018). Ioannidis (2005) provides a probabilistic argument for why most published research is false. Franco, Malhotra, and Simonovits (2014) assess that, among a sample of high-quality awarded research proposals in social sciences, the majority of those that find null results are never written, and the likelihood of publication increased dramatically with the strength of the results. Gerber and Malhotra (2008a; 2008b) showed that, in top psychology and political science journals, the number of papers with p-values just below 0.05 were more frequent than those just above 0.05 by a factor of two and three, respectively. (Brodeur et al. 2016) found similar irregular behavior for the distribution of p-values in top economic journals.

Lack of Reproducibility

Parallel to this work and in similar fashion, multiple disciplines have begun to assess issues of replicability and reproducibility of previously published research. Replicability tests whether or not the same results could be obtained in a different setting (data) using the same procedures (methodology) (LeBel et al. 2017). Reproducibility tests if it is possible for a third party to obtain the same results using the same data, methods, and code (Goodman et al., 2016). Replicability has been part of the scientific method for centuries, and reproducibility has become increasingly important with the predominance of computation in empirical work. A large-scale replication effort attempted to replicate the results of 100 studies in psychology; using several criteria, this effort found approximately 33-50% of the original findings to be observed in the replication study (Open Science Collaboration 2015). In a similar exercise for behavioral economics, 11 studies were replicated out of a total of 18 (Camerer et al. 2016). Regarding reproducibility (same data, code, and methods), Stodden, Leisch, and Peng (2014) describe the importance of improving current standards for computational science. In an exercise to assess the

reproducibility of 67 papers in macroeconomics, (Chang and Li 2015) were able to obtain qualitatively similar results for only 29 papers, while six papers could not provide proprietary data. More recently, Gertler, Galiani, and Romero (2018) attempted to re-run the analysis code from a sample of 203 empirical papers from leading journals in economics and was able to obtain the same results in only 14% of the papers. These issues are a subset of what some authors refer to as the reproducibility crisis (Baker 2016), or the credibility crisis (Stodden, 2014), in science. This crisis has had the positive effect of bringing to forefront the core scientific principles of replication, reproduction, and openness (Merton 1942), though it has also shed light on how the scientific community simultaneously accepts these principles but does not practice them on a regular basis (Anderson, Martinson, and De Vries 2007). Fortunately, a strong response to the credibility crisis in science has begun to emerge, under the umbrella of open science.

THE OPEN SCIENCE MOVEMENT

Open science is "the practice of science in such a way that others can collaborate and contribute, where research data, lab notes and other research processes are freely available, under terms that enable reuse, redistribution, and reproduction of the research and its underlying data and methods" (Bueno de la Fuente 2017). Miguel et al. (2014) argue for three guiding norms to promote transparency in empirical social sciences: (i) disclosure of key details involved in the analysis and collection of the data; (ii) registration of pre-analysis plans that contain information on the outcome variable, independent variable(s) of interest, model specifications, and other analytic choices before the data is collected; and (iii) open access to data, code, and additional documentation. To help journals, funders, and scholarly societies apply these norms, Nosek et al. (2015) developed the Transparency and Openness Promotion (TOP) Guidelines. The TOP

Guidelines contain eight standards, and compliance with each standard is defined on four different levels. Level 0 (lowest) represents the status quo, where journal policy does not mention the open science practice. Level 1 is achieved when authors of the research disclose the extent to which they follow the standard. Level 2 is attained when the journal requires adherence to a given standard. Level 3 (highest) is met when the journal enforces adherence to a given standard (e.g., reproduced reported analyses independently prior to publication). For further detail on how to achieve the highest level and follow best practices, (Christensen, Freese, and Miguel 2019)provide a manual for best practices in research transparency, and Kitzes, Turek, and Deniz (2017) present a set of 31 case studies of computational reproducible research.

APPENDIX 2: THE PROCESS OF POLICY ANALYSIS

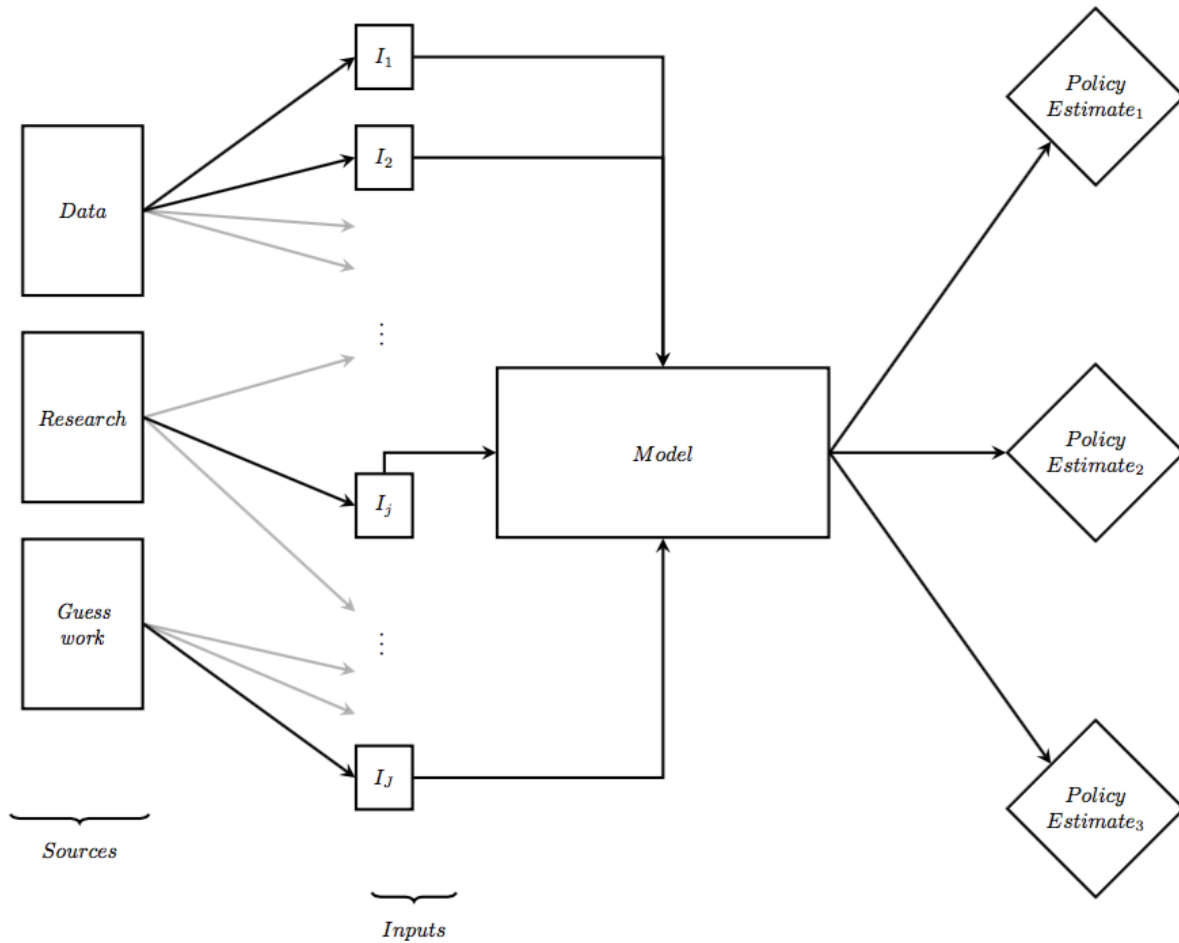**Figure A1. The Process of Policy Analysis**

Figure A1 represents a simplified model of the process involved in a policy analysis. Analysts use three primary sources for inputs into the analyses: information from previous research (e.g. elasticities, behavioral parameters), data to contextualize the specific policy issue (e.g. micro-data for the specific context where the policy issue is discussed), and guesswork to fill in any missing pieces required to complete the analysis (e.g. extrapolation parameters, take-up rates, distributional effects). All of these sources are used to generate inputs that are used in a model (e.g. micro-simulation, cost benefit analysis), and this model produces the policy estimates to be used by policymakers.