**Staff Member A**

The PIs have submitted a thorough and well-written proposal. I do wonder what, if any, might be the primary vulnerabilities that would prevent the PIs from making the headway on the three projects that they have identified.  What would skeptics say of the methodological choices?

* For questions #1 & 2:
  + Our method depends on our matched researchers being a good control group. There is, of course, no way to verify this *a priori*, so at this point the best we can do is discuss steps we are taking to make the control group as valid as possible. Our matching strategy will consider both similarity of researcher characteristics (e.g. years of experience, previous publication quality) and similarity on research topics. As noted in our technical appendix (p. 43): “The advantage of using machine learning for matching, rather than traditional matching methods such as propensity score matching and nearest neighbor matching, is that it allows us to evaluate the quality of the overlap for the specific features above. For example, we can ensure that our matched pairs both have similar research interests relevant for the grant, and be more lenient on both having similar years of experience.”   
      
    Note that the machine learning begins with input from the research team, as we must first create training data. For a subset of researchers, we will use topic matching to identify a potential comparison group of researchers working on similar topics. The research team will manually identify those researchers we view as most similar to our funded researchers. This will enable us to ensure that relevant characteristics are considered. Moreover, to improve validity, we will have multiple research team members evaluate each match.
  + Should we find that machine learning does not produce quality matches, we will consider other matching strategies, such as Coarsened Exact Matching (CEM) as a backup plan. Should no matching strategy provide a quality control group, we would hve to temper our presentation of results by noting that causal inferences are not possible, and that our results merely present correlations. However, given the researchers’ previous experience with matching research portfolios, we believe that this is unlikely.
  + The key assumption of matching is that there are no unobserved differences among researchers that influence the choice of research project. By using matched controls in a difference-in-difference setting, we are able to use individual researcher fixed effects to control for any remaining unobserved differences after matching.
* For question #3
  + Here we ask whether the distance between a scientist’s previous work and a new publication affect the impact and novelty of the publication. Endogeneity is less of a concern here: our measures of novelty compares a publication’s similarity to the previous literature. This measure may be sensitive to the number of prior years considered, and we will do sensitivity analysis on this window to evaluate that possibility. Our measure of impact depends on citations received. Since these citations are made after the article is published, they should be endogenous to the research in the publication itself.

**Staff Member B**

Ambitious and important set of research questions, as well as what appears to be a well-thought out research design. My only question has to do with the possible self-selection bias(es) of those who seek out the federal funding versus those who don't. Could they be younger, less well-funded, or more willing to take risks in funding?

* Good question: matching is designed to incorporate these biases – that is, our matching focuses not only on similarity of research topics, but also research characteristics Thus, while it may be true, for example, that younger researchers are more likely to seek out federal funding, our matching allows us to compare researchers at similar stages of their careers.
* In addition, our regressions allow us to test whether researcher characteristics matter (e.g p. 16). While the difference-in-difference model in our main specification has the advantage of controlling for any unobserved differences across researchers, it doesn’t allow us to test whether the effects of funding are different for different types of researchers. “Thus, we will also estimate a richer model that includes various researcher controls in place of these fixed effects. For the treatment effect in such a model to be causal, there must be no additional unobserved variables that affect research choice. A comparison between the results from these models and the difference-in-differences specification allows us to verify whether our controls are sufficient.”

**Staff Member C**

Interesting approach to understanding the impact of funding calls on research. The use of machine learning is relatively straightforward - in fact my one suggestion is that they might be a little more creative with the application of those tools. ML tends to perform better when leveraging known structures in the data, and scientific articles are highly structured. In addition to the proposed topic modeling of the full text of articles in the sample, could they also break the articles into standard sections (introduction, lit review, methods, discussion) and run a similar distance analysis? I would imagine that the methods may vary wildly while other sections will converge more rapidly in response to funding, an effect that I assume might be the desired output but would be muted by studying only the abstract or full text. Would love to hear what they think on that point.

**Staff Member D**

Proposal seems reasonable but I defer to experts on methodology. I am surprised that the government has not conducted it's own evaluation. Maybe I missed it, but what is the universe of $$/grants/researchers included in the ARPA-E, EERE, and relevant NSF programs? How many publications are associated with this universe?

* This is a good question, although one that we can only answer in part until we begin the research.
* To give a sense of the funds involved, the figure below, taken from Popp (2019 – my Hamilton brief), shows the R&D budget for ARPA-E and major renewable energy technologies from 2010-2018. Total spending in these programs ranges from 1.6 to 2.2 billion dollars. However, it is important to note that only a subset of this spending will be represented in our data, as our focus is on funds given to academic researchers. For the two ARPA-E programs used as an example in our proposal, 10 of the 13 Electrofuel awards went to academics, whereas for GENI, 6 of 15 awards went to academics. We anticipate a greater share of awards to go to academics from the DOE Office of Science compared to EERE.  
    
  As for publication counts, we identified XXXX publications attributed to the 16 researchers in our sample in Scopus.  
    
  For further evidence of the number of publications available, Goldstein and Narayanamurti (2018) identify publications acknowledging awards made by ARPA-E, the DOE Office of Science, and EERE between fiscal years 2009 and 2015. They identify over $3 billion of awards, although this includes awards made to for-profit entities. They identified 5,181 papers acknowledging these awards by the end of 2016. For our purposes, note that these are just papers acknowledging a DOE award. Our sample will include all papers attributed to the primary investigators, which will be a much larger sample.
* To identify relevant NSF programs, we will start with the pool of researchers funded by the DOE programs in our sample and look for NSF programs that that also support these researchers. We expect that many relevant programs will come from the Engineering and Mathematical and Physical Sciences directorates.

**Staff Member E**

While it seems straightforward to determine whether new researchers have entered a field, it’s less clear why this may have happened. A new funding opportunity is certainly an important factor but it’s hardly the only factor, and possibly not even the most important factor. Researchers are strongly influenced by how interested they are in a given question and whether the research raises or lowers their academic status. I suppose the more general point is that it’s not just the appearance of a new funding opportunity but also the specific details of the opportunity that determine whether a researcher will change his/her direction. How will the proposers control for these and similar factors? In spite of these complications, understanding what’s changed (researchers change direction in response to a new opportunity) is valuable even if the underlying motivations are not crystal clear.

* We control for other factors that may influence researchers in two ways. I elaborate on both and discuss their strengths and weaknesses below.
  + Matching and DiD – doesn’t allow us to identify factors
  + Control variables (e.g. policy, prices as in Popp 2017). Allows us to identify reasons, but doesn’t control for omitted variables – perhaps put this first.
  + Note not a concern for the third research question – there “new” to a field is exogenous to those citing the research in the future

**Staff Member F**

A very important research question that is methodologically challenging to answer. As such, there are lots of technical details surrounding the construction of the control group. I always feel like a bit of faith is necessary when using synthetic controls. One key assumption is that the treatment cannot have any impact on the outcome variables for the control sample of researchers. It isn’t hard to imagine, for example, that when a Funder announces the availability of funds for a certain topic, not only the eventual winners end up focusing some of their energy into this topic. Even failed proposals require thinking hard enough to produce at least a sensible project outline. Who is to say some of these aren’t eventually completed? Perhaps even systematically. I don’t say this to take anything from their empirical approach. It’s solid and I’m sure they’ve discussed ways to tackle these sorts of issues. But it is also complex.

* This is a nice point. Our response would be that finding funding elsewhere is part of what we care about – that is, if researchers can find other sources of funding, then the funding call didn’t change the direction of research. Staff member F’s point challenges this assumption: the idea here is that failed projects get funded elsewhere, and that failed projects were done *in response to a specific funding call*. If so, that would bias our results downward – funding would appear less effective.
  + Will require some care when comparing targeted/open programs, as this endogeneity concern should only affect targeted programs.

**Staff Member G**

The set of questions proposed, which boil down to evaluating the impact of the funding strategies used by agencies to advance certain areas of research, are very interesting and policy-relevant. If we can learn something reliable here, it can (and hopefully will) be used to craft future government programs to make them more effective. Beyond the funding and publication data itself, the proposed work relies centrally on a number of elements: (i) a measure of “similarity” and “novelty” based on machine learning applied to a textual analysis of abstracts and Funding Opportunity Announcements; (ii) a similar measure of a researcher’s “breadth,” and (iii) a complex matching procedure for identifying a control group of “untreated” scientists. I will be particularly interested in to see what external reviewer’s opinion are of the significance and validity of these complex measures, since all the conclusions will depend on them.

You will see that Staff Member C has some suggestions on the Machine Learning portion of the project, Staff Member D has some questions about overall scope, ***and A, E, & F have methodological questions – these last ones mainly in the spirit of really wanting to understand more about the methods and how the synthetic control groups will be formed and less so that they have fundamental concerns about them. I would be particularly sure to address these questions.***

**Response to Reviewers and Staff Members**

**Does Government Funding Change What You Do? The Effects of Funding on the Direction and Impact of Academic Energy Research**

**Principal Investigators: David Popp and Daniel Acuña**

**Response to Methodological Questions: Regression Analysis**

Staff Member A (SM A) asks what skeptics might say about our methodological choices. Some possibilities are raised by other external reviewers, such as: self-selection of those seeking funding, (SM B), other factors determining research interests (SM E), and the influence of funding opportunities on non-funded researchers (SM F). Below we discuss threats to the validity of our results, including those raised by reviewers, and how we plan to address them. We divide the discussion into concerns addressing research questions #1 and #2, which are similar, and research question #3, which uses a different data set and different methodology.

*Research questions #1 and 2:*

For research questions #1 and 2, which identify the effect of funding on choice of topic, our method depends on our matched researchers being a good control group. These controls are intended to address concerns raised by SM B and SM E. For example, our matching focuses not only on similarity of research topics, but also research characteristics. While it may be true, as raised by SM B, that younger researchers are more likely to seek out federal funding, our matching allows us to compare researchers at similar stages of their careers. Moreover, our regressions allow us to test whether these researcher characteristics matter (e.g. p. 16 of our proposal). While the difference-in-difference model in our main specification has the advantage of controlling for any unobserved differences across researchers, it doesn’t allow us to test whether the effects of funding are different for different types of researchers. It is for that reason that we also plan to estimate a richer model that includes various researcher controls in place of these fixed effects. For the treatment effect in such a model to be causal, there must be no additional unobserved variables that affect research choice. A comparison between the results from these models and the difference-in-differences specification allows us to verify whether our controls are sufficient. If they are, regressions with researcher characteristics will help us determine what types of researchers are most affected by different funding opportunities.

Similarly, researchers may be influenced by external factors that make research on a topic more interesting or provide more professional reward (IE E). While matching controls for differences across researchers, our solution to this concern is through the difference-in-difference (DiD) regression. Our matched group of funded researchers and controls deals with concerns about motivations that different types of researchers might have, as described above. In contrast, here we are concerned with external changes that affect *all* researchers. For instance, increasing concerns about climate change may make working on clean energy research more attractive. Our DiD framework assumes that these external forces influence similar funded and non-funded researchers in the same way. We will test this by checking that our matched treated and control researchers have similar publication trends *prior to the funding date* (e.g. similar pre-trends). Thus, the treatment we observe from funding is whether receiving funding causes the funded researcher to deviate from pre-funding publication trends more than her matched control. DiD regression also helps alleviate any remaining concerns from matching. The key assumption of matching is that there are no unobserved differences among researchers that influence the choice of research project. By using matched controls in a difference-in-difference setting, we are able to use individual researcher fixed effects to control for remaining unobserved researcher attributes after matching.

SM F raises an important caveat that our matching and DiD regressions may not fully address. This reviewer notes that new funding calls will attract proposals that are not selected for funding, but that are good enough to receive funding elsewhere. Consider an example where we observe no impact of funding on research direction – both funded and non-funded researchers do more energy research over time. Our interpretation using the modeling strategy above would be that it wasn’t the funding that increased energy research, but rather external factors that increased the viability of energy research for all scientists. SM F’s point challenges this assumption. Failed applications may get funded elsewhere, and these failed projects were done in response to a specific funding call. If so, that would bias our results downward – funding would appear less effective.

As failed applications are not part of the public record, we cannot test this alternative interpretation directly. However, we can look for possible signs suggestive of this in our results. For instance, we will estimate year-by-year impacts for funding. Because it will take additional time for failed projects to find funding elsewhere, a scenario such as proposed by SM F would suggest that funding leads to an immediate movement towards energy research, with the effect disappearing over time as failed applications are funded elsewhere and completed at later dates. Comparing results between targeted and open funding programs can also shed light on this, as the concern raised by SM F should only be an issue for targeted funding programs that solicit proposals on pre-defined topics.

*Research question #3*

In research question #3 we ask whether the distance between a scientist’s previous work and a new publication affects the impact and novelty of the publication. That is, does bringing new researchers into a field generate high impact research? Endogeneity is less of a concern here. Our measure of novelty focuses on the similarity between a publication and the previous literature, which is fixed at the time of publication. The measure of similarity may be sensitive to the number of prior years considered, and we will do sensitivity analysis on this window to evaluate that possibility. Our measure of impact depends on citations received. Since these citations are made after the article is published, they should be exogenous to the research in the publication itself. What does matter is opportunities for future citations. Papers will receive more citations if there are more subsequent papers available to cite them. In our citation regressions, we control for factors such as energy prices and policy likely to affect the popularity of clean energy as a research field.

**Response to Methodological Questions: Machine Learning and Matching Methods**

There is, of course, no way to verify the quality of our matching *a priori*, so at this point the best we can do is discuss steps we are taking to make the control group as valid as possible. Our matching strategy will consider both similarity of researcher characteristics (e.g. years of experience, previous publication quality) and similarity on research topics. As noted in our technical appendix (p. 43): “The advantage of using machine learning for matching, rather than traditional matching methods such as propensity score matching and nearest neighbor matching, is that it allows us to evaluate the quality of the overlap for the specific features above. For example, we can ensure that our matched pairs both have similar research interests relevant for the grant, and be more lenient on both having similar years of experience.”

Note that the machine learning begins with input from the research team, as we must first create training data. For a subset of researchers, we will use topic matching to identify a potential comparison group of researchers working on similar topics. The research team will manually identify those researchers we view as most similar to our funded researchers. This will enable us to ensure that relevant characteristics are considered. Moreover, to improve validity, we will have multiple research team members evaluate each match.

Should we find that machine learning does not produce quality matches, we will consider other matching strategies, such as Coarsened Exact Matching (CEM) as a backup plan. Should no matching strategy provide a quality control group, we would hve to temper our presentation of results by noting that causal inferences are not possible, and that our results merely present correlations. However, given the researchers’ previous experience with matching research portfolios, we believe that this is unlikely.

***need a response from Daniel on machine learning techniques***

**Response to Questions about Scope and Impact**

SM D asks about the universe of grants and researchers included in the programs we propose to study. This is a good question, although one that we can only answer in part until we begin the research. To give a sense of the funds involved, the figure below, taken from Popp (2019), shows the Department of Energy R&D budget for ARPA-E and major renewable energy technologies from 2010-2018. Total spending in these programs ranges from 1.6 to 2.2 billion dollars. However, it is important to note that only a subset of this spending will be represented in our data, as our focus is on funds given to academic researchers. For the two ARPA-E programs used as an example in our proposal, 10 of the 13 Electrofuel awards went to academics, whereas for GENI, 6 of 15 awards went to academics. We anticipate a greater share of awards to go to academics from the DOE Office of Science compared to EERE.

As for publication counts, we identified XXXX publications attributed to the 16 researchers in our sample in Scopus. This includes their complete publication record, not just publications directly generated by these ARPA-E projects. For further evidence of the number of publications available, Goldstein and Narayanamurti (2018) identify publications acknowledging awards made by ARPA-E, the DOE Office of Science, and EERE between fiscal years 2009 and 2015. They identify over $3 billion of awards, although this includes awards made to for-profit entities. They identified 5,181 papers acknowledging these awards by the end of 2016. For our purposes, note that these are just papers acknowledging a DOE award. Our sample will include all papers attributed to the primary investigators, which will be a much larger sample.

We can say less about the relevant NSF programs, as those will be identified during our research. To decide which NSF programs to include, we will start with the pool of researchers funded by the DOE programs in our sample and look for NSF programs that that also support these researchers. We expect that many relevant programs will come from the Engineering and Mathematical and Physical Sciences directorates. Methods Appendix 1 provides examples of NSF programs that funded selected researchers funded by ARPA-E’s Electrofuels program.

Finally, note that the sample for research question #3 will be different, as it will focus on the universe of publications in biofuels, energy storage, solar energy, and wind energy from 1991-2017. These data will extend a data set collected for work by the PI in Popp (2016, 2017). Those data, complete through 2011, include over 70,000 publications.

**References**

Goldstein, A.P. and Narayanamurti, V. 2018. Simultaneous pursuit of discovery and invention in the US Department of Energy. *Reserch Policy* 47(8): 1505-1512.

Popp, D. 2019. “Promoting Innovation for Low-Carbon Technologies,” Policy Proposal 2019-14, The Hamilton Project, Brookings Institution, Washington, DC.

Popp, D. 2017. From Science to Technology: The Value of Knowledge From Different Energy Research Institutions. *Research Policy* 46(9): 1580-1594.

Popp, D. 2016. Economic analysis of scientific publications and implications for energy research and development. Nature Energy 1(4): 1–8.