**Syracuse Reviewer #3**

1. **What is your view as to the qualifications of the PIs? How likely is it that the PIs will be able to accomplish their goals in the timeframe outlined?**

The researchers propose combining machine learning and regression analysis using RFPs and publications. Specifically, they propose to use “machine learning (ML) analysis of text to track changes in research portfolios and measure novelty”. They use ML to produce the data to analyse. This also includes using ML to create a counterfactual of ‘unfunded’ researchers. Researchers propose using regression analysis to understand the relationship between different funding types and changes in the overall ‘portfolio’ of researchers, separate from general trends. The PIs have complementary expertise. Prof. Popp has demonstrated expertise in regression analysis and Prof. Acuna on machine learning of text. Overall, they have very good qualifications and demonstrated experience.

The proposed project has a duration of 3 years and they propose producing 3 papers. One paper would compare the extent to which the papers produced funded by targeted ARPA-E solicitations are different in topic from those published by ‘non-ARPA-E funded’ researchers that were similar; one on whether ‘new’ researchers (meaning first-time recipients of ARPA-E funding?) produce more impactful research when compared to similar researchers funded through other open calls (the researchers point to NSF and BES); and one on whether ML provides something different from matching. Again, producing three papers for this budget and timeline seem doable.

I am including here a short review of the questions and approach. It took me quite a while to get a handle on the research questions that they will actually try to answer and what the comparisons/counterfactuals would be. After rereading several sections, this is my understanding of the research questions (with some reflections/questions):

1. Whether or not targeted R&D funding opportunities in ARPA-E result in a change in the focus of the research of scientists compared to the evolution of similar scientists that are not funded by ARPA-E. Here one question is whether they know (and whether it matters) if they have been funded in other ways. Can they exclude researchers funded by EERE, DoD, NSF?;
2. the extent to which different types of calls are associated with different changes in the direction of research of awardees. They propose this by comparing the change in the research of targeted ARPA-E and EERE awardees compared to those that have been awarded open ARPA-E, BES and NSF programs (note that this is clearer in the page 8 but unclear before that). The discussion about Q2 in page 2 indicates that the goal is to understand whether these different funding opportunities attract ‘new’ researchers but the discussion that follows indicates that this question is not about which funding instrument attracts ‘new’ researchers, but rather whether it changes their research. These seem to be different things. New researchers could be defined in other ways: e.g., researchers that were in energy, but in a different part; researchers that were in the science but not in the technology; researchers that were not applying to anything and then applied to ARPA-E; researchers that had never applied to ARPA-E, etc. I understand that (a) and (b ) are using the same data for the outcome variable (change in the direction of research using topic modelling) but just using different comparison groups.
3. whether researchers that changed their direction (I am not clear about whether here the researchers are just using the subsample of recipients of targeted ARPA-E) funding are ‘more’ successful (measured by the production of more novel and highly cited papers) compared to other papers in the same (new for the researcher) field. Again, here it would have bene useful to know whether they are just going to calculate the level of difference of research for ARPA-E awardees, or just take a general sample of energy publications, extract the authors, and compare the ‘success’ of the articles that had been written by people that have previously published different articles with those in the sample of energy articles.

Overall, I would have liked to see the proposal indicate in the first page the cases they are looking at, since I tis mainly ARPA-E. I would argue that because ARPA-E is so different it may be difficult to extract insights for other funding sources. At least the authors should be upfront about this and discuss the possible limitations to external validity. In the first page of the actual proposal the researchers indicate that they cover both ARPA-E and NSF. Then Q1 only includes ARPA-E. I am a little unclear about what RFPs are covered in what questions, and I would have liked to see an explanation of why those agencies were chosen.

As I will explain in the strengths/weaknesses part, I am worried about selection, omitted variable bias and interpretation of the results of the change in direction. How are they going to test the impact of other funding design variables, such as funding size, process for review, people reviewing it, agency awarding it, etc.?

I also think the proposal would benefit from more focus on the question about how to produce more policy-relevant research. Depending on how they operationalize the ‘new researcher’ category, the information can become easier or harder to interpret. How would one ‘attract’ new researchers if that is just measured by changes in topic models of publications? Have the authors thought about trying to code departments, or training, or to give more thought about this ‘new researcher’ interpretation? To use the example provided, materials science has been integral to solar panels for a long time. And it is not clear that, even if it was more fringe, one could conclude from their independent variable (researcher that changed publications) that policy should target material science researchers (or any other topic/area of science/technology). Or what a ‘novel researcher’ means (if it means running topic models). In page 5 the authors say ‘the collaborations [that will be studied here will consider] researchers coming from different scientific communities’. **Again, the metric of change in direction may not have anything to do with different scientific communities**, depending on how the topic modelling is done. This will need validation.

1. **What will be the value of the proposed research work to the field? How significant, unique, or worthwhile a contribution would this project make?**

The most important contribution that this paper can make will be two-fold:

1. To understand whether or not different types of R&D funding approaches (open and targeted) result in different changes in research papers, compared to a relevant comparison group
2. To determine whether those researchers that changed published ‘better’ papers than other papers in that field

I agree with the authors that there has not been as much analysis in cleantech on researcher impact and incentives as in health. Understanding the impact of open vs targeted solicitations in cleantech and the relationship to publication success would be very useful. So I think that, with some additions, clarifications and additional checks/variables, this proposal could fill an important scholarly and policy gap.

1. **What are the strengths and weaknesses of the project and the proposed work plan?**

However, to make it as influential as possible, in addition to the policy-relevance question, the authors should better address **selection bias, omitted variable bias** (comparing ARPA-E with other funding programs that differ from it), and the **limitations of the outcome metrics** for ‘publication success’.

On the point about **selection bias**, this proposal is doing something different from what Azoulay et al (2011) did. While Azoulay et al focused on the extent to which the more long-term incentives and flexibility provided by HHMI were associated with more creative outcomes compared to NIH, controlling for the quality of the researchers. In this proposal, the main outcome variable for questions 1 and 2 is change of topics in publications and the main outcome variable in question 3 is about the impact of the change in direction in the quality of the research. But there is no discussion of how to control for the quality of the researchers (which Azoulay et al 2011 tried to do using a set of early career prizes). They do not discuss the possible impact of selection bias. It could be the case that, because ARPA-E is so selective, they just filter in the best people and any improved outcomes in the quality of what they do compared to the average researcher in their topic is not due to their change in focus (the spillover) but rather because they are better. I would have liked to have read more on selection bias. Using ML to find researchers in a similar area defined by topic modelling and with similar demographic characteristics (page 3 of the proposal) indicate that they will not try to control for researcher quality. This is crucial. I am not sure that the matching approach as currently described can be described as ‘enabling us to provide a more thorough causal analysis’ (page 4). In page 16 researchers talk about some researcher controls (not specified, beyond accounting for ‘junior researchers’), but quality is not included.

I would also have liked to see more discussion in Q2 about what **other factors may yield changes** in direction beyond **open/targeted.** In other words, the extent to which comparing ARPA-E with researchers without funding or with open programs may suffer omitted variable bias. For example, the size of funding, the duration, the review process may prove to be important mechanisms that are needed (beyond or in addition to targeted solicitation) to change the direction of the researcher.

And why not use more comprehensive **metrics of publication success**? Have the novelty and citation measures been validated in this field or any field? How do they link to ‘success’? They could use interviews, complement them with some metric of the extent to which the publications contributed to a commercialized technology, etc. I know this is not easy, but it would be good to at least discuss the limitations of these novelty and citation metrics.

A minor question is related to when the researchers indicate they plan to compare the texts of “publications and requests for proposals of funded scientists compared to the evolution of the research of a control group with similar research portfolios prior to the funding decision”. Here I am a little confused of how they are going to use both publications and RFPs in the treatment group and just publications in the control group. At least they should propose testing whether results hold when just comparing publications in treatment and control groups, using a combination of differences in differences plus matching. I do not think the researchers mention this combination of methods explicitly, so it would be good if they could clarify whether this is the plan they have for Q1.

Other relevant papers that are not cited include:

1. Stephan, T.S. Schmidt, C.R. Bening, V.H. Hoffmann (2017): „ The sectoral configuration of technological innovation systems: Patterns of knowledge development and diffusion in the lithium-​ion battery technology in Japan.” *Research Policy*, 46, 709-​723. [doi: 10.1016/j.respol.2017.01.009](http://www.sciencedirect.com/science/article/pii/S0048733317300082)
2. Stephan, A., Bening, C.R., Schmidt, T.S., Schwarz, M., Hoffmann, V.H. (2019): "The role of inter-​​sectoral knowledge spillovers in technological innovations: The case of lithium-​​ion batteries" *Technological Forecasting and Social Change*, 148, 119718. [doi: 10.1016/j.techfore.2019.119718](https://doi.org/10.1016/j.techfore.2019.119718" \t "_blank)

There are many strengths, I will highlight a few:

* discussion of the datasets and how they will link them was strong
* discussion of how to deal with multiple teams of authors

1. **What is your evaluation of the appropriateness of the budget? Are there any elements that you would add, eliminate, or change?**

One PhD studentship per stream including summer funding over 3 years seems appropriate. The PIs are requesting one month per year each, which also seems appropriate.

1. **What additional suggestions or recommendations for improvement, if any, do you have on the overall project that might make the proposal stronger?**

I have already made several suggestions in response to previous questions.