National Science Foundation

NO-DEADLINES SYNTHETIC CONTROL AND EXPLORATORY OUTCOMES ANALYSIS AND RECOMMENDATIONS FOR A FUTURE RIGOROUS

May 2022

A report from the Evaluation and Assessment Capability Section of the National Science Foundation.

About the Evaluation and Assessment Capability Section

The Evaluation and Assessment Capability (EAC) Section bolsters NSF efforts to make informed decisions and promote a culture of evidence. Located in the Office of Integrative Activities of the Office of the Director, EAC provides centralized technical support, tools, and resources to conduct evidence-building activities and to build capacity for evidence generation and use across the agency. EAC is led by NSF's Chief Evaluation Officer.

About this report

This report was prepared for EAC under contract number 49100421F0221. The views expressed are those of the authors and should not be attributed to NSF, nor does mention of trade names, commercial products, or organizations imply endorsement of same by the U.S. Government.

Preferred citation

Fesler, Lily and Lindsay Fox. 2022. *No-deadlines Synthetic Control and Exploratory Outcomes Analysis and Recommendations for a Future Rigorous Evaluation* [Memorandum]. Alexandria, VA: National Science Foundation.

NSF Quality Certification: Level 2

Quality Certifications

Level 1 — The author(s)/contractor(s) are responsible for the quality and conclusions presented in this report

Level 2 — NSF verified that this report underwent quality assurance procedures and contributed to assessing its content

Level 3 — NSF independently reproduced the analysis presented in this report

Memo

To:	Taylor Rhodes and Robyne McRey
From:	Lily Fesler and Lindsay Fox
Date:	3/23/2022
Subject:	No-deadlines synthetic control and exploratory outcomes analysis and recommendations for a future rigorous evaluation

Purpose

This memo explores the options for using synthetic control methods to examine the effects of no-deadlines (NDL) approaches using National Science Foundation (NSF) administrative data.

Summary

We find that a synthetic control method is a promising approach for examining the impacts of NDL for programs with enough appropriate counterfactual programs (in other words, programs in directorates in which only a subset of programs switched to NDL and for which a weighted average of outcomes for non-NDL programs looks like NDL programs in the years before NDL implementation). We conduct an exploratory outcomes analysis of two NDL programs in the Directorate for Geosciences (GEO) (out of 12 NDL programs in our sample) and find that NDL substantially reduced the number of proposals received (by about 120 to 130 proposals). We also see some evidence that NDL may have increased the amount of requested funding, decreased the number of reviewers, and decreased the percentage of proposals rated by reviewers to be fair or underperforming for these two NDL programs. These findings are exploratory and not generalizable to other programs or directorates.

We also discuss how considering different analytic decisions and resolving certain data issues could improve a future rigorous evaluation of an NDL approach. For example, future research is needed to determine the advantages and disadvantages of conducting the analysis at different time intervals (for example, annual versus quarterly), how to treat programs that switch to NDL in the middle of a fiscal year, and the extent to which principal investigators (PIs) may anticipate NDL. This analysis could also be improved by better understanding potential data quality issues for variables used in determining NDL status, improving program linkages across time, more closely examining programs that implement NDL for only a subset of proposals, and refining the process for choosing comparison programs.

The remainder of this memo is organized as follows. We present (1) the goal of the synthetic control analysis, (2) our approach for identifying comparison programs for NSF programs that have implemented an NDL approach, (3) findings from an exploratory outcomes analysis, and (4) considerations for a future rigorous evaluation of an NDL approach.

Goal of the synthetic control analysis

Ideally, to estimate the impact of using an NDL approach, we could conduct a randomized controlled trial (RCT). In an RCT, we would randomly assign some NSF programs to implement NDL and assign others to not implement NDL in a given year. An RCT would ensure that NDL and non-NDL programs do not differ on any characteristics other than their implementation of NDL, and thus any observed differences between NDL and non-NDL programs could be directly attributed to the NDL approach. However, programs at NSF have made their own decisions as to whether they implement NDL, and thus, conducting an RCT is not feasible in this scenario.

Instead, we attempt to approximate an RCT as closely as possible by using a synthetic control approach. In this approach, we identify non-NDL programs that are as similar as possible to NDL programs in the years before NDL implementation. Once we establish similarity at baseline, we can compare outcomes between NDL and non-NDL programs to estimate the impact of NDL. A synthetic control approach is preferable to a simple pre-post analysis if other factors affecting outcomes are present around the time of implementation. For example, in fiscal year (FY) 2019, a five-week government shutdown forced NSF to stop holding review panels, which might have decreased the overall number of proposals (National Science Board 2020). For programs that switched to NDL in FY 2019, it would be important to understand how much of any decline in proposals might be attributed to the implementation of NDL versus the impacts of the government shutdown.

Programs and directorates have chosen to implement NDL at different times in the last ten years. They varied in their motivations and goals for implementing the no-deadlines approach, including reducing NSF staff and reviewer workload, increasing funding rates, distributing workload throughout the year, and providing greater flexibility to PIs to think creatively, build interdisciplinary teams, and submit better proposals (Patino and Garcia 2020; Yuan et al. 2020; Lane 2021). We focus our synthetic control analysis and exploratory analysis on the following outcomes that may change as a result as a result of switching to NDL: (1) proposal volume, (2) proposal quality, (3) burden on reviewers, (4) level of collaboration on each project, and (5) amount of funding requested.[1]

Approach for identifying comparison programs for NDL programs

Identifying a comparison group is a two-step process. First, we identify NDL and non-NDL programs in each fiscal year, then we create a comparison group by weighting non-NDL programs to look as similar as possible to NDL programs before NDL implementation.

Process for identifying NDL and non-NDL programs

To identify NDL and non-NDL programs, we begin with all programs that received any new proposals in response to funding opportunities from FY 2009 to FY 2021.[2] We then link programs' funding opportunities across fiscal years so we can measure program characteristics across time.[3] Lastly, we identify the programs that implemented an NDL approach in each fiscal year using the listed close dates in the proposal data. Proposals responding to a funding opportunity with a deadline should have a close date listed in the proposal data, and proposals responding to a funding opportunity without a deadline

[1] We also considered funding rate and dwell time as potential outcomes but, based on guidance from NSF, did not ultimately include them.

[2] Following National Science Foundation (2021), we exclude supplements, forward funds, renewals, principal investigator transfers, and other non-new projects. We also exclude proposals that respond to the Proposal & Award Policies & Procedures Guide (PAPPG) instead of a specific program, proposals that are not in RPTSQL (such as contracts), and proposals submitted in a fiscal year different from the funding opportunity deadline. NSF received 781,835 proposals for 921 programs from FY 2009 to FY 2021, and after these exclusions we include 648,748 proposals for 836 programs in our analytic sample. See the technical appendix for more details on these exclusions.

[3] We link funding opportunities using a data set that NSF provided to us, as well as NSF program announcement publications. See the technical appendix for more details on this process.

should have no close date listed in the proposal data. If most proposals for a given program have no close dates in a given fiscal year, we identify that program to be using an NDL approach in that fiscal year.[4]

Estimating counterfactual outcomes for NDL programs

For our analysis, we use the augmented synthetic control method to estimate counterfactual outcomes for NDL programs, had NDL not been implemented (Ben-Michael, Feller, and Rothstein 2021a). Athey and Imbens (2017) have called synthetic controls "arguably the most important innovation in the policy evaluation literature in the last 15 years," as they use a specifically weighted average of comparison units for which the pre-implementation outcome trends match those of the treated units more closely than a simple average across all or a selected subset of comparison units. The outcomes of this weighted group of comparison units over post-implementation periods are then used to estimate the counterfactual outcomes of units subject to the policy implemented during that time, in absence of the policy. Augmented synthetic controls can further improve pre-treatment fit by allowing for a broader range of weights than traditional synthetic controls (Ben-Michael et al. 2021a). By directly targeting reducing differences in pre-trends, this approach focuses on outcome trajectories over time. As such, it is likely to yield a more appropriate counterfactual comparison in this context than approaches that focus on matching average outcome levels between programs for one or more time periods, such as matching on propensity score or Mahalanobis distance (Abadie 2021).[5]

To ensure the group of programs that contribute to estimating counterfactuals for each NDL program are plausibly comparable, we require these programs to be in the same directorate as the NDL program.[6] This ensures that any directorate-wide changes happening in a given year are reflected in both the treatment programs and their counterfactuals. For example, the Directorate for Biological Sciences eliminated its preliminary proposals for some core programs at the same time it switched to an NDL approach (National Science Foundation 2017). Estimating counterfactuals for programs in other directorates that did not eliminate preliminary proposals in that same year might be problematic.[7]

Synthetic control methods are also dependent on having enough pre- and post-treatment data for each cohort, as well as enough comparison units (Abadie 2021). To ensure we have enough pre- and post-treatment data, we balance our data by only including programs that had at least six years of data before the NDL implementation year (for fitting synthetic control weights) and at least three years of data after

[4] We are not aware of a systematic data source that indicates whether each funding opportunity or program is implementing NDL by fiscal year. Proposal close dates are generally consistent with the information in NSF program announcement publications (see the technical appendix for more details on this).

[5] Even though propensity score matching under a multi-level framework can help construct more appropriate comparison groups for specific programs, propensity score models do not specifically match on outcome trends over time and thus typically do not produce matches that are as strong in settings with panel data.

[6] We define programs to be in a directorate if at least 50 percent of proposals that program receives are submitted to that directorate. Further, we tried matching across directorates but, after discussions with NSF, concluded that this led to poorer matches.

[7] In several interviews, we asked program officers, program directors, and division directors if any programs were similar to the programs they are connected with that never adopted an NDL approach, or adopted an NDL approach at a different time, that could be used as a comparison in a study. One interviewee suggested using programs in their directorate, one suggested a division outside of their directorate with similar funding rates, and a third did not suggest another program or division, but simply suggested doing an experiment to test the effects of an NDL approach. The suggestion of the first interviewee to use within-directorate programs is consistent with other guidance we have received from NSF for this matching analysis and informs our methodological approach.

NDL implementation for both NDL and comparison programs.[8] For example, for NDL programs that first implemented in FY 2019, we only include NDL and comparison programs with data from FY 2013 to 2021 and match on the pre-implementation years of FY 2013 to FY 2018. Because our analysis is conducted at the fiscal-year level, we also remove NDL programs that switched to an NDL approach in the middle of a fiscal year.[9] We include programs as potential comparisons if they do not implement NDL over the included time period (six years pre-implementation and three years post-implementation).[10]

To generate a unique set of synthetic control weights for each NDL program and each outcome, we focus on five outcome variables: number of proposals received, average requested amount for the project, average number of collaborative proposals per project, average number of reviewers per proposal, and percentage of proposals rated exceptional, fair, and underperformed. The primary variables used to generate weights for each outcome are the values of that outcome in each pre-intervention year, but we also include the other outcomes as additional covariates to use for generating weights (as the outcomes might be related to one another).[11] We use ridge regression to generate the weights for the control units to construct the synthetic control (Ben-Michael et al. 2021a).[12]

Findings from an exploratory outcomes analysis

In collaboration with NSF, we selected two programs in GEO to show in an exploratory outcome analysis: Antarctic Research (NSF 21-567) and Arctic Research Opportunities (NSF 21-526).[13] We find that these two NDL programs received 120 to 130 fewer total proposals relative to the counterfactual. Figures 1 and 2 show the trends pre- and post-implementation for the five primary outcomes.

First, we examine fit for the pre-treatment years to assess the quality of the match. We assess fit in two ways: by examining the pre-implementation trends and by assessing model fit metrics. For a strong match, we would expect the synthetic control (that is, the estimated counterfactual) to track the NDL program very closely in all pre-treatment years. For example, in Figure 1, we look for close fit from FY 2009 to FY 2017 (the year before implementation began) for all outcomes. In both Figures 1 and 2, we

---

[8] In choosing the number of years to include in the analysis, we need to balance two competing interests: (1) to include enough years of data that we are matching on longer-term trends (as opposed to fluctuations or noise in the data in a given year), and (2) to not require so many years of data that we need to exclude NDL and comparison programs unnecessarily. Requiring six years of pre-implementation data allows us to examine the synthetic control fit without dramatically reducing the number of programs that can be included in the analysis. Requiring three years of post-implementation data will allow us to examine any short-term impacts and whether any impacts increase or dissipate over time (Abadie 2021).
[9] This removed 15 NDL programs. We conduct our matching analysis at the fiscal-year level because many programs’ deadlines are annual, not quarterly.
[10] See Tables A7–A9 in the technical appendix for the programs included in each directorate.
[11] For example, the model specification when the outcome variable is the number of proposals includes covariates for average requested amount; average number of collaborative proposals per project; average number of reviewers; and the average percentage of proposals rated as exceptional, fair, and underperforming.
[12] The regularization parameter in the model penalizes the distance from traditional synthetic control weights, which reduces extrapolation bias. See the technical appendix for the ridge-augmented synthetic control estimator.
[13] We chose these programs after examining the pre-implementation trends for the 12 NDL programs in our analytic sample. The pre-implementation trends for the other 10 NDL programs are included in the technical appendix. Post-implementation years were not included in the graphics until we chose the 2 NDL programs for the exploratory outcomes analysis.

see a close match for all pre-treatment years, indicating that the synthetic control approach was able to construct comparison groups that trend similarly to the NDL programs.[14]

Second, we examine the post-implementation years to assess how outcomes for two NDL programs might change after implementation, relative to their counterfactual. Figures 1 and 2 suggest that relative to the counterfactual, these two NDL programs received fewer proposals in the three years after NDL implementation. Changes for the other outcomes are less visible in these figures. Model estimates suggest that NDL programs received 123 to 128 fewer proposals, requested $265,000 to $294,000 more funding, required 0.9 to 1.8 fewer reviewers, and were less likely to be rated as fair (by 4 to 9 percentage points) or underperforming (by one percentage point).[15] There were not consistent patterns for number of collaborative proposals per project and exceptional reviews. There is conflicting evidence as to whether differences may increase or decrease after the first year of NDL implementation across the two programs.[16] These figures and findings are exploratory and are not generalizable to NDL programs more generally.

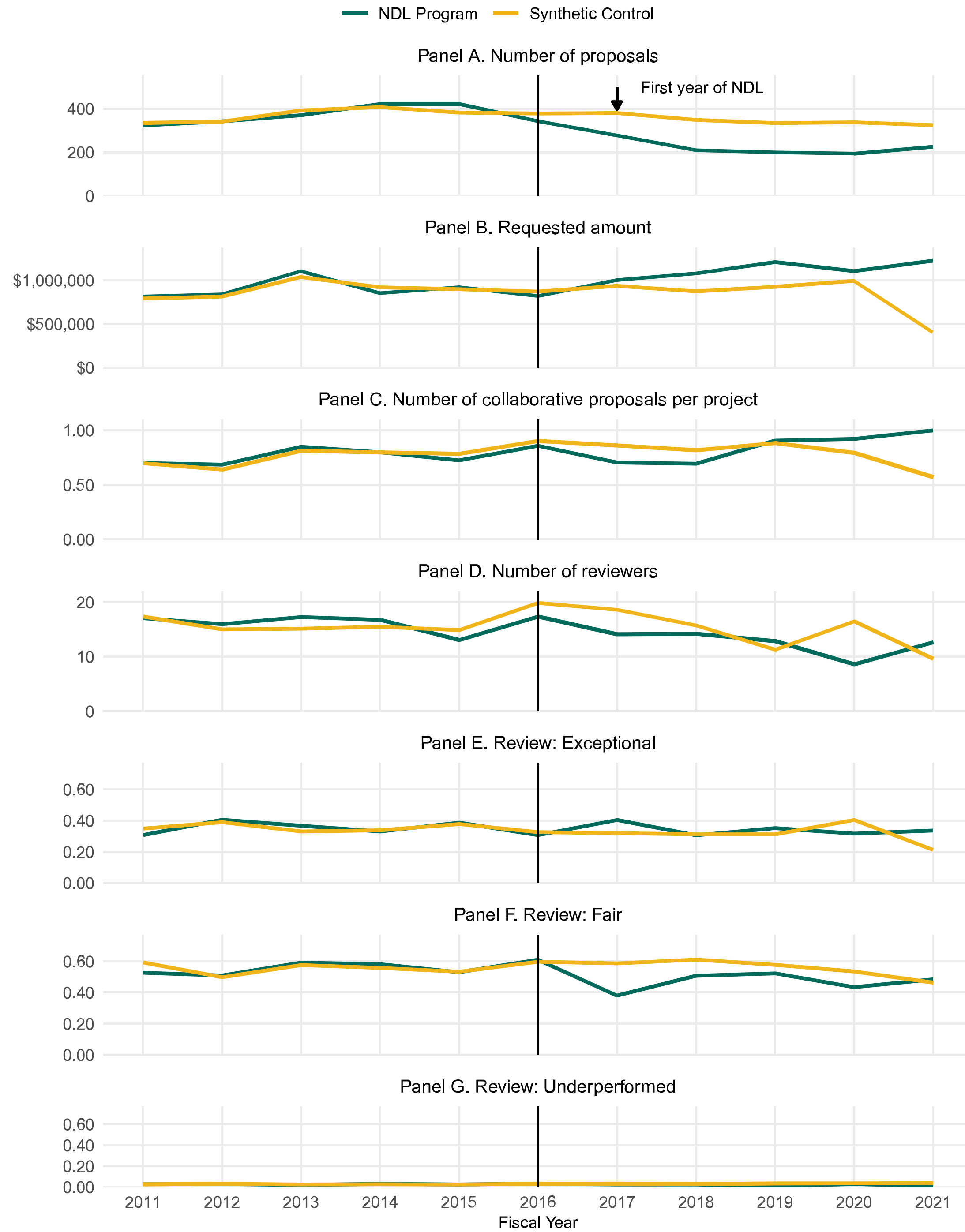
The synthetic control approach performed relatively poorly for some other programs. For example, the pre-implementation trends for the Division of Integrative Organismal Systems Core NDL Program (NSF 21-506) did not vary substantially from year to year, but its estimated counterfactual exhibits some sharp increases and decreases in the pre-implementation years (see Figure A10 in the technical appendix). This is likely because there are very few (only four) potential comparison programs for BIO, making it difficult to estimate counterfactuals that closely align their outcomes over the pre-implementation years. Indeed, BIO programs generally achieved the worst fit statistics as well (see Table A10 in the technical appendix).

[14] The L^2 fit statistic is also relatively low for those two programs, which indicates a better match (see Table A11 in the technical appendix).
[15] See Table A12 for all estimated magnitudes. We do not present any tests of statistical significance.
[16] Differences increased in magnitude after the first year for a few of the outcomes (including number of proposals, requested amount, and underperformed proposals for NSF 21-526) and decreased in magnitude for all other outcomes (see Table A12).

Figure 1. NDL and synthetic control trends before and after NDL implementation for GEO: Antarctic Research (NSF 21−567)

Note: The synthetic controls are weighted averages of the programs in GEO that have at least six years of data before NDL implementation, at least three years of data after NDL implementation, and did not implement NDL during this period. This includes seven comparison programs (see Table A8). The weights differ by outcome. There are 1,562 projects in the NDL program and 7,152 projects across the programs in the synthetic control prior to NDL implementation.

Figure 2. NDL and synthetic control trends before and after NDL implementation for GEO: Arctic Research Opportunities (NSF 21−526)



Note: The synthetic controls are weighted averages of the programs in GEO that have at least six years of data before NDL implementation, at least three years of data after NDL implementation, and did not implement NDL during this period. This includes six comparison programs (see Table A8). The weights differ by outcome. There are 1,261 projects in the NDL program and 4,030 projects across the programs in the synthetic control prior to NDL implementation.

Considerations for a future rigorous evaluation

To inform a future evaluation of the NDL approach, we discuss five areas for consideration: (1) potential enhancements of our use of synthetic controls, (2) other analytic decisions to consider for a future evaluation, (3) data issues that could improve the analyses, (4) advantages and disadvantages of conducting an RCT, and (5) other outcomes to consider for a future analysis.

Potential enhancements of our use of synthetic controls

Synthetic control methods perform best when many pre-intervention years and many relevant comparison units are available. Our data include six to nine pre-intervention years and three to seven comparison programs, both of which are smaller than in typical synthetic control studies.[17] This presents two risks: (1) that we would not be able to construct counterfactuals that perform well and (2) that the counterfactual would be overfit to the pre-intervention years (in which case the synthetic control program would not be a good counterfactual for the treatment program in the post-intervention years). We do find that the synthetic control methods are not able to produce strong counterfactuals for some NDL programs (for example, some BIO NDL programs) due to having few relevant comparison units, and thus we did not select those programs for the exploratory outcomes analysis. A future analysis could consider whether some programs beyond those in each directorate could serve as potential comparison programs.[18] We also help avoid overfitting by using regularization via ridge regression.

The fit of synthetic control models can also be improved by including covariates that are predictive of post-intervention outcome values (Abadie 2021). We included the outcomes that were the focus of our analysis as covariates in our synthetic control estimation, but a future evaluation should examine the most predictive covariates more deeply.[19]

Synthetic control models can also be strengthened by investigating impacts across multiple units that might be exposed to treatment at different times (Ben-Michael, Feller and Rothstein 2021b). This can improve the power of the model and reduce the chances that any observed impacts are due to circumstances specific to one treatment–comparison combination or one particular time period. A future analysis could use synthetic control methods to estimate the impacts of NDL across multiple programs. For example, a future evaluation could analyze programs from GEO, ENG, and BIO in a single analysis to generate a cross-directorate average impact of NDL programs.

Other analytic decisions to consider for a future evaluation

Our analysis is conducted at the fiscal year, which is the most straightforward way to conduct an analysis because many programs have annual deadlines. Conducting an analysis at the biannual or quarterly level instead is possible, particularly if some directorates commonly had more frequent deadlines before implementing an NDL approach. The advantage of conducting an analysis at a more frequent interval than annual is twofold. First, it would double or quadruple the number of time points, which would

---

[17] The issue of relatively few pre-intervention years could potentially be ameliorated by two issues discussed below: by changing the time interval from annual to more frequently (for example, biannual or quarterly) and by expanding the data to include more years before FY 2009. However, it is also possible that neither of those approaches would be feasible.

[18] We found that including all programs across directorates as potential comparison programs led to poor matches. However, a future analysis could examine whether programs that are cross-listed in that directorate could serve as appropriate comparisons.

[19] As one example, dwell time may be predictive of the number of reviewers.

strengthen the synthetic control design. Second, it would allow us to include NDL programs in our analysis that began using NDL partway through a fiscal year (instead of toward the beginning or end of a fiscal year). However, conducting an analysis at a more frequent interval might not be possible, particularly if very few proposals are commonly received in a given time interval (for example, NSF might receive very few proposals in December for a June proposal deadline). A future feasibility study should examine the extent to which deadlines occur multiple times per year across NDL and potential comparison programs to assess whether increasing the time interval would be possible.

Relatedly, we exclude NDL programs from our analysis that switch to NDL mid-year. For the purposes of this exploratory memo, we exclude programs that received 10 percent or more proposals that were not subject to deadlines in the year before implementation. A future evaluation should examine the sensitivity of the analyses to different thresholds, including lower thresholds (such as 1 percent) and higher thresholds (such as 20 percent). A future evaluation could also consider removing the implementation year from the analysis and comparing post-implementation years to pre-implementation years for programs that implement NDL mid-year.

A future feasibility study should also examine the extent to which principal investigators might anticipate NDL before the implementation year and alter their proposal decisions before implementation begins (Abadie 2021). For example, directorates commonly announce their decision to switch to NDL in a Dear Colleague letter, which may be distributed in a fiscal year before the implementation year. It is possible that principal investigators alter their decisions to submit a proposal in the fiscal year before implementation if they know the subsequent fiscal year will not have deadlines. This would make comparisons between the prior fiscal year and the first implementation year more challenging. A future feasibility study should examine the timing of the Dear Colleagues letters relative to the NDL implementation year and further investigate NDL programs in which anticipation is more likely.

Data issues that could improve the analyses

There are several data issues that, if resolved, would make it easier to conduct a future evaluation. First, we use the close date associated with each proposal to determine whether a proposal is subject to deadlines. Proposals without associated close dates are very commonly proposals submitted in response to an NDL program (and frequently match the deadline information included in NSF program announcement publications). However, the close date information does not always match that in the NSF program announcement publications. It would be helpful to understand the extent to which this is due to data entry errors versus real variations in deadlines. Collecting these data more systematically would improve our process for identifying NDL and non-NDL programs.

Second, our analysis relies on being able to track programs across time. Because programs have different program numbers over time, we need to be able to link these program numbers to include them in our analysis. For this exploratory analysis, we used a data set that NSF provided to us and created our own linking system based on NSF program announcement publication data, but we do not believe the linkages we have are comprehensive. If NSF could construct a comprehensive data set that links NSF program announcement publications over time, we would be able to include more programs in our analysis. Similarly, sometimes a single program replaces multiple programs. For our analysis, we treat these previous programs as precedents to the current program. It would be helpful to understand whether that is the best way to link programs across time.

Third, some proposals within certain programs are subject to NDL and others are not subject to NDL in a given fiscal year. For example, Computer and Information Science and Engineering (CISE) Core Programs are implementing NDL for small projects but not for medium projects or Office of Advanced Cyberinfrastructure (OAC) Core Projects. It would take additional work and collaboration with NSF to understand the appropriate proposals with which to compare these NDL programs before NDL implementation. For example, for CISE, we would need to accurately identify OAC proposals before and after implementation, as well as small and medium projects (and whether definitions as to which projects may be small versus medium might change over time).

Fourth, further institutional knowledge may be needed to identify appropriate potential comparison programs, particularly when unique program characteristics are not observable in available data. For example, we learned from NSF that two programs in the Directorate for Engineering (ENG) (Industry-University Cooperative Research Centers Program and Spectrum and Wireless Innovation enabled by Future Technologies) that were included as potential comparison programs would likely not serve as appropriate counterfactuals for the ENG NDL programs because of their unique characteristics. Further collaboration with NSF would be helpful to continue refining the list of potential comparison programs in each directorate.

Fifth, not all funding opportunities in our data have associated program names.[20] If we had a more comprehensive dataset that links all program opportunities with program names, we would be able to include more programs in our analysis.

Sixth, fewer proposals are in the data before FY 2009 than after, so we began our analysis with FY 2009.[21] However, if data quality is strong in and before FY 2008, then including additional pre-implementation years could strengthen our analysis.

Advantages and disadvantages of conducting an RCT

Instead of conducting a synthetic control analysis to approximate an RCT, another option would be to conduct an RCT going forward. For example, NSF could randomly select a subset of programs that have not yet implemented NDL to begin implementing NDL in FY 2023 and track outcomes for both NDL and non-NDL programs for the next several years. This would eliminate any concerns about bias or unobservable differences between NDL and non-NDL programs. However, this would also require NSF to wait several years to learn about the impacts of NDL and may require randomizing a relatively large number of NSF programs to implement NDL. Programs may also prefer to choose whether and when to implement NDL, so it may not be as feasible as other designs.

[20] For example, we do not observe program names in NSF databases for PD 16-014, PD 18-6881, or PD 22-514. We need program names to contextualize each program and understand 1) whether it could serve as a counterfactual program (for non-NDL programs) or 2) which programs would be the strongest counterfactual programs (for NDL programs). For example, if we did not know that NSF 20-570 represented Industry-University Cooperative Research Centers Program, we would not have known that NSF 20-570 may not be an appropriate counterfactual for the ENG NDL programs.

[21] Between FY 2009 and FY 2021, there are between 53,000 and 67,000 proposals per year in Solr. In FY 2007 and FY 2008, there are substantially fewer proposals (about 30,000 and 40,000, respectively). There are very few (less than 3 proposals per year) before FY 2007. The increase between FY 2008 and FY 2009 does not seem to be due to an actual large increase in proposals between those two years (National Science Foundation 2010). We thus assume that the difference in number of proposals in Solr between FY 2008 and FY 2009 is due to a difference in the types of proposals that were collected in Solr before FY 2009 and exclude prior years.

Other outcomes to consider for a future analysis

A future evaluation should consider which outcomes are of primary interest based on the logic model of the NDL approach and which can be measured accurately and reliably. For example, validated administrative data related to the diversity of NSF's portfolio might be valuable to include in an analysis. Other outcomes, such as those related to NSF staff and reviewer workload and processes, may need to be created and collected systematically to use in an analysis.

References

Abadie, Alberto. 2021. "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects." *Journal of Economic Literature* 59(2):391-425.

Athey, Susan and Guido W. Imbens. 2017. "The State of Applied Econometrics: Causality and Policy Evaluation." *Journal of Economic Perspectives* 31(2):3-32.

Ben-Michael, Eli, Avi Feller, and Jesse Rothstein. 2021a. "The Augmented Synthetic Control Method." *Journal of the American Statistical Association* 116(536):1789-1803.

Ben-Michael, Eli, Avi Feller, and Jesse Rothstein. 2021b. "Synthetic Controls with Staggered Adoption" (Working Paper 28886). Cambridge, MA: National Bureau of Economic Research.

Lane, Paul. 2021. "Removal of Submission Deadlines from TMRPs." Presented to the National Science Foundation, February 17, 2021.

National Science Board. 2020. *National Science Foundation's Merit Review Process: Fiscal Year 2019 Digest*. Retrieved January 20, 2022 (https://www.nsf.gov/nsb/publications/2020/merit_review/FY-2019/nsb202038.pdf).

National Science Foundation. 2010. *FY 2009 Performance and Financial Highlights*. Retrieved February 22, 2022 (https://www.nsf.gov/pubs/2010/nsf10002/nsf10002.pdf).

National Science Foundation. 2017. "Dear Colleague Letter: Implementation of 'No-Deadline,' Full-proposal Submission Process for Most Programs in the Directorate for Biological Sciences" Retrieved February 22, 2022 (https://www.nsf.gov/pubs/2018/nsf18011/nsf18011.jsp).

National Science Foundation. 2021. "Querying Proposal and Award Data Resources and Data Sources." Prepared for New Employee Roundtable, October 2021.

Pankow, Neha. 2020. *Removal of Deadlines for Programs across NSF: A Template for Case Studies*. Washington, DC: National Science Foundation.

Patino, Lina and Diana Hernandez Garcia. 2020. "Removing Proposal Submission Deadlines in Earth Sciences." Presented to the National Science Foundation, December 2, 2020.

Yuan, Grace, Gabriela Niño de Guzmán, Yuen Lau, Steven Zehnder, Richard Nash, Jesus Alvelo Maurosa, and Miriam Scheiber. 2020. "ENG and No Deadlines." Presented to the National Science Foundation, December 3, 2020.

Technical Appendix

I. Introduction

The no-deadlines (NDL) synthetic control and exploratory outcomes analysis is designed to explore options for a future evaluation of the NDL approach. This technical appendix accompanies the main memo and describes the data, sample, and methods used in greater detail.

II. Data

In this section, we describe the data used in the analysis. We collected data from the National Science Foundation's (NSF) Solr Application Programming Interface (API) and other NSF databases, and NSF program announcement publications. We then assembled these data into an analytic data set that we use for the synthetic control and exploratory outcomes analyses.

A. NSF data systems

We used NSF's Solr API, NSF's Fastlane database, and NSF's Report Server to collect proposal-level information for NSF programs. We included all programs that received proposals from fiscal year (FY) 2009 to FY 2021. We collected information on variables including proposal status, proposal received date, requested amount, and number of reviewers. See Table A1 for a full list of variables that we collected from Solr and other NSF databases.

Table A1. Data elements, descriptions, and sources

Data element	Description	NSF source
Id	Proposal ID	Solr
prop_id	Proposal ID	FLflpdb.flp.prop, rptdb.csd.prop, parsdb.csd.rev_prop, flpdb.csd.rev_prop
award_amount	Award amount	Solr
Directorate	NSF directorate	Solr
Division	NSF division	Solr
lead_id	Lead proposal ID	Solr
lead_proposal	Lead proposal flag	Solr
natr_rqst_code	Nature of request code	Solr
program_announcement	Program announcement	Solr
Received	Proposal received date	Solr
received_year	Proposal received year	Solr
requested_amount	Requested amount	Solr
reviewer_count	Reviewer count	Solr
Status	Proposal status	Solr
pgm_annc_id	Program announcement ID	FLflpdb.flp.prop_covr
clos_date	Close date on proposal cover	FLflpdb.flp.prop_covr
pgm_annc_titl	Program announcement title	FLflpdb.flp.pgm_annc
fund_type	Program announcement type	FLflpdb.flp.pgm_annc

Data element	Description	NSF source
revr_id	Reviewer ID	parsdb.csd.rev_prop, flpdb.csd.rev_prop
rev_prop_rtng_code	Reviewer rating code (for proposal quality)	flpdb.csd.rev_prop ratings

B. NSF program announcement publications

We also collected information about solicitations and program announcements from publicly available NSF publications. For all program solicitations and announcements included in the Solr proposal data from FY 2009 to FY 2021, we searched for the relevant NSF publication via the URL format: https://www.nsf.gov/pubs/[program fiscal year]/[program solicitation number]/[program solicitation number].htm. From each publication, we collected program title, the previous program solicitation or announcement, and program deadline information. For NDL programs, the deadline information states, "Proposals Accepted Anytime." We identified publication information for 71 percent of funding opportunities.

C. Assembling the database for analysis

Step 1. Identifying the population of NSF funding opportunities

We identified the population of programs as those that received any proposals in response to funding opportunities from FY 2009 to FY 2021. We merged NSF's Solr data with proposal-level data in other NSF databases to construct a data set with all variables of interest.

Step 2. Linking funding opportunities across time within programs

Second, we linked funding opportunities over time because program numbers change over time. For example, the Arctic Research Opportunities program's current program solicitation number is NSF 21-526 (which is an NDL solicitation), and its previous solicitation numbers have included NSF 16-595 (also an NDL solicitation), NSF 14-584 (a non-NDL solicitation), and others. We linked these solicitation numbers together so we could measure characteristics of the Arctic Research Opportunities program both before and after NDL implementation. We began by using the data set that NSF provided to us on November 22, 2021, that linked many program solicitations, announcements, and descriptions. For program solicitations and announcements that do not have linkages in these data, we used the linkages that we developed from the NSF program announcement publication data. For program descriptions in the data set that were missing some linkages, we assumed that program descriptions with the same digits after the hyphen were linked (for example, PD 13-1517 and PD 16-1517). If linkages were not included in the NSF data set and we were not able to construct them ourselves using program announcements and solicitations, these programs will not be linked in our data.

Step 3. Identifying the programs that implemented an NDL approach in each fiscal year

Third, we identified the programs that implemented an NDL approach in each fiscal year using the listed close dates in the proposal data. Proposals responding to a funding opportunity with a deadline should have a close date listed in the proposal data, and proposals responding to a funding opportunity with NDL should have no close date listed in the proposal data. If 50 percent or more proposals for a given program have no close dates in a given fiscal year, we identify that program to be using an NDL approach in that fiscal year.

We also validated the proposal close dates by comparing them with the information we found in NSF program announcement publications for program solicitations and program announcements (

Table A2). The vast majority (93 percent) of program solicitations and announcements that listed having NDL on the NSF program announcement publications received almost no (less than 5 percent) proposals with close dates listed, and a similar percentage (93 percent) of program solicitations and announcements that listed having deadlines on the NSF program announcement publications received almost all (more than 95 percent) proposals with listed close dates. This indicates that close dates accurately represent NDL status for the vast majority of proposals, although there may be some discrepancies in the data.

Proposal close dates may not perfectly map to the information in NSF program announcement publications for two reasons: (1) data entry errors, and (2) proposals submitted toward the beginning or end of a fiscal year that are submitted in response to a previous or later solicitation with a different NDL status. We can only collect information on deadlines for program descriptions via the close date (and not via NSF program announcement publications), so we cannot do this comparison for program descriptions.

Table A2. Percentage of projects with close dates, by whether NSF program announcement publications list deadlines for all projects, deadlines for some projects, or no deadlines for all projects

Projects with close dates	Deadlines for all projects (percentage)	Deadlines for some projects (percentage)	No deadlines for all projects (percentage)
0 to less than or equal to 5 percent	0.8	25.0	92.7
More than 5 to less than or equal to 25 percent	0.0	0.0	7.3
More than 25 to less than or equal to 50 percent	0.2	4.2	0.0
More than 50 to less than or equal to 75 percent	0.8	0.0	0.0
More than 75 to less than or equal to 95 percent	5.2	25.0	0.0
More than 95 to less than or equal to 100 percent	92.9	45.8	0.0
Total	**100.0**	**100.0**	**100.0**

Note: Close date refers to whether the data show that proposals contain a due date. NSF program announcement publications are listed as having deadlines if all proposals are subject to deadlines, mixed deadlines if a subset of proposals are subject to deadlines, and no deadlines if no proposals are subject to deadlines.

D. Constructing variables of interest

We constructed five key variables of interest:

- **Number of proposals.** This includes all proposals except those that have been withdrawn, returned, or deleted.[22] It includes preliminary proposals and proposals with award decisions still pending.[23]

[22] Fewer than 0.01% of received proposals are deleted.
[23] Withdrawn and returned proposals are typically excluded from most analyses (National Science Foundation 2021). However, previous NDL reports have also highlighted these proposals as key variables of interest (Pankow 2020). A future evaluation should consider whether to include these variables as outcomes of interest.

- **Average requested amount.** This represents the average amount of funding that each project requested.
- **Average number of collaborative proposals per project.** This is calculated as the number of collaborative proposals per project (or lead proposal).
- **Average number of reviewers.** This includes all reviewers who are associated with a proposal in the data set and represents the reviewer pool. This includes both reviewers who agreed and declined to review the proposal.
- **Average proposal quality.** We follow ENG and report three proposal quality categories (Yuan et al. 2020):
  - **Exceptional:** proposal has an average review score of 4 or above
  - **Fair:** proposal has an average review score below 4 and a maximum review score of 3 or above
  - **Underperformed:** proposal has a maximum review score of below 3[24]

III. Sample

We restricted our analysis to new projects, excluding supplements, forward funds, renewals, and principal investigator (PI) transfers (using the "Nature of Request" variable).[25] We removed proposals that were submitted to the Proposal & Award Policies & Procedures Guide (PAPPG) (for example, NSF 20-1) because these cannot be tied to a specific solicitation.[26] We also removed proposals that were received before or after the fiscal year in which the proposal was due.[27] Table A3 shows the number of proposals and programs that were removed by each exclusion criterion. After applying these proposal exclusions, our sample included 648,748 proposals and 836 programs (out of 781,835 proposals received for 921 programs during our time period). This is 83 percent of proposals received and 91 percent of programs that were active between FY 2009 and FY 2021.

To prepare for our synthetic control analysis, we also applied some program-level exclusions. We excluded programs without a title in the database, programs without at least six consecutive years of data, and NDL programs that did not switch to NDL at the end of a fiscal year.[28] Our analytic data set included 468,561 proposals across 194 programs. This is 60 percent of all proposals and 21 percent of all programs. Many programs do not have six consecutive years of data, so this analysis will be more representative of programs that are held consistently across years.

[24] Since observed proposal rating is a proxy for proposal quality, future analysis may examine other proposal quality methodologies used by Yuan et al. (2020) such as analyzing maximum review score—where at least one reviewer scores an E (Excellent), V (Very Good), or E/V.

[25] This follows guidance from the National Science Foundation (2021) which states that "most queries should be filtered for Nature of Request = New Project." It is also consistent with the exclusion criteria used by ENG and Earth Sciences (Patino and Hernandez Garcia 2020; Yuan et al. 2020).

[26] According to the National Science Foundation (2021), proposals may be submitted to the PAPPG instead of the solicitation, and thus funding opportunity searches may not include all proposals.

[27] NDL status is determined by fiscal year. Thus, proposals that NSF receives in a fiscal year different from the fiscal year in which the solicitation is due might be subject to a different deadline from other proposals received in that year.

[28] As discussed in the memo, we exclude programs without a title in the database because we need program titles to understand whether and how to include the program in the synthetic control analyses.

Table A3. Number of proposals and programs by exclusion criteria

	Number of proposals	Number of programs
Total number FY 2009 – FY 2021	**781,835**	**921**
Proposal exclusions		
Supplements, forward funds, renewals, PI transfers, and other non-new projects	83,679	65
Proposals that respond to the PAPPG instead of a specific program	37,784	14
Proposals that are in Solr but not in RPTSQL	96	1
Proposals submitted before or after the close date fiscal year	11,528	5
Total number included after proposal exclusions	**648,748**	**836**
Programs with no title in the database	16,197	43
Programs without 6 consecutive years of data	141,014	584
NDL programs that did not switch to NDL at the end of a fiscal year	22,976	15
Total number included in analysis	**468,561**	**194**

Table A4 shows the characteristics of programs included in our analytic data set by fiscal year. The average program in our data set receives 222 proposals in support of 191 projects per year requesting $977,057 in funding (and receiving an average of $799,545 in funding). Thirty-six percent of funding opportunities are via program descriptions and 43 percent are via program solicitations. The majority of proposals (55 percent) receive a fair rating, 27 percent receive an exceptional rating, and 5 percent are reviewed as underperforming. Twenty-eight percent of proposals are funded, and the average time to decision is 172 days. A small minority of proposals (0.4 to 0.7 percent) are responding to NSF's Rapid Response Research (RAPID) and EArly-concept Grants for Exploratory Research (EAGER) funding mechanisms.

Table A5 shows that Engineering (ENG) had the most programs included in this analysis over this time period (22 percent of all programs), followed by Mathematical & Physical Sciences (MPS) (18 percent of all programs) and Geosciences (GEO) (16 percent).

Table A4. Characteristics of programs by fiscal year

Characteristics	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	Full period
Average number of proposals	218.6	239.0	216.6	221.5	225.4	234.8	248.2	227.8	239.3	210.4	206.6	201.6	192.1	222.4
Average number of projects	192.0	210.1	188.0	194.8	197.1	204.4	213.5	197.1	203.2	179.6	171.7	169.0	158.3	191.3
Project average requested amount ($)	968,214	934,032	922,092	756,199	841,052	1,011,671	850,278	838,383	814,201	816,675	1,997,889	1,018,099	1,010,891	977,057
Project average award amount ($)	813,497	848,635	844,709	671,226	696,773	737,598	805,326	805,229	782,765	726,149	991,659	979,655	716,019	799,545
Number of collaborative proposals per project	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.3
Type of funding opportunity (%)														
Program description	23.8	24.0	30.9	33.6	35.4	41.9	41.6	38.3	38.9	38.3	36.0	37.7	37.9	35.6
Program announcement	1.6	1.3	1.3	1.2	1.7	1.1	1.1	0.6	0.6	0.6	0.0	0.0	0.0	0.9
Program solicitation	1.5	2.2	9.3	18.5	33.5	55.1	56.1	60.5	60.0	60.4	64.0	62.3	62.1	42.8
Proposal quality (%)														
Exceptional	29.5	29.2	28.8	26.7	25.8	26.8	26.5	27.9	26.3	26.5	26.0	28.4	26.9	27.3
Fair	53.2	53.7	55.2	55.0	52.6	55.6	53.4	55.0	55.8	55.6	57.2	55.1	54.0	54.7
Underperforming	4.7	4.2	4.8	5.1	5.3	5.4	5.2	5.4	6.2	5.1	5.5	4.7	4.3	5.1
Average time to decision (days)	187.8	173.5	173.1	176.2	179.4	172.5	163.9	178.6	172.7	171.4	171.8	169.2	149.4	172.3
Funded (%)	30.8	27.3	26.3	26.9	26.6	25.9	28.8	26.5	26.3	28.2	30.8	32.8	33.7	28.4
Cross-directorate (%)	9.7	10.0	10.2	10.7	11.0	10.3	11.2	11.6	11.8	11.8	11.5	12.5	11.3	11.1
Proposal type (%)														
RAPID	0.1	0.4	0.3	0.3	0.1	0.2	0.6	0.4	0.3	0.2	0.2	1.4	0.8	0.4
EAGER	0.8	1.1	0.8	0.5	1.4	0.4	1.0	0.3	0.3	0.5	0.9	0.6	0.8	0.7
Number of programs	134	150	157	168	173	184	178	173	170	161	156	152	151	2,107

Table A5. Percentage of programs in each directorate by fiscal year

Directorate	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	Full period
Biological Sciences (BIO)	8.9	9.7	9.4	8.7	9.0	8.0	8.3	9.0	8.7	8.0	9.5	9.0	9.0	8.8
Computer & Information Science & Engineering (CISE)	4.8	5.3	5.4	6.6	6.8	6.7	6.8	7.3	7.1	8.2	7.8	7.8	7.9	6.8
Education & Human Resources (EHR)	11.3	10.1	9.9	8.6	9.0	9.6	9.0	10.0	10.5	9.7	10.1	10.5	11.3	9.9
Engineering (ENG)	18.7	23.2	24.6	23.2	23.7	23.7	23.1	19.8	21.6	21.9	21.0	21.5	20.2	22.1
Geosciences (GEO)	20.4	17.8	17.6	16.3	15.8	15.4	15.3	15.8	14.8	16.3	16.2	15.4	15.5	16.3
Mathematical & Physical Sciences (MPS)	18.3	17.6	17.5	20.4	20.4	19.1	18.6	18.5	18.1	16.4	14.8	16.1	17.4	18.0
Office of the Director (O/D)	3.0	2.8	2.0	2.5	2.1	2.2	2.3	2.3	1.8	1.9	1.9	1.9	2.0	2.2
Social, Behavioral & Economic Sciences (SBE)	14.7	13.5	13.6	13.7	13.2	15.3	16.6	17.2	17.3	17.7	18.6	17.7	16.8	15.8
Number of programs	134	150	157	168	173	184	178	173	170	161	156	152	151	2,107

Table A6 shows the percentage of programs using an NDL approach by fiscal year. Over half to three-quarters of programs in the ENG, Biological Sciences (BIO), and GEO directorates were using an NDL approach by FY 2021, compared to under 15 percent of programs for the MPS, Computer & Information Science & Engineering, and Social, Behavioral & Economic Sciences directorates.

Table A6. Percentage of programs using an NDL approach, by fiscal year

Directorate	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Geosciences	33.3	30.8	33.3	33.3	37.0	39.3	40.7	33.3	40.0	42.3	52.0	56.5	56.5
Mathematical & Physical Sciences	12.5	8.0	7.7	11.8	8.8	11.8	6.2	9.7	10.0	12.0	13.6	13.0	12.0
Biological Sciences	9.1	7.1	7.1	7.1	6.7	7.1	21.4	6.7	14.3	50.0	57.1	61.5	61.5
Engineering	0	14.7	2.6	5.3	5.0	4.7	5.1	5.9	10.8	17.1	51.5	74.2	74.2
Computer & Information Science & Engineering	0	0	0	0	0	0	0	0	0	0	9.1	9.1	0
Social, Behavioral & Economic Sciences	5.3	10.0	9.5	13.0	4.5	7.1	10.0	10.3	6.9	7.1	6.9	7.4	8.0
Education & Human Resources	0	0	0	0	0	0	0	0	0	0	0	0	0
Office of the Director	50.0	50.0	0	25.0	0	50.0	50.0	50.0	66.7	33.3	33.3	33.3	33.3
Number of programs	134	150	157	168	173	184	178	173	170	161	156	152	151

Note: We define programs as being in a particular directorate if at least 50 percent of proposals submitted to the program are for that directorate. Programs in which no directorate receives at least 50 percent of proposals are not included in this table.

IV. Synthetic Control Methods

A. NDL and comparison programs included in synthetic control analysis

Twelve NDL programs in our sample have enough pre- and post-implementation years of data. The synthetic control group for each of these comparisons is drawn from the program's directorate (as discussed in the memo). Our sample includes 2 NDL programs in GEO, 5 programs in ENG, and 5 programs in BIO (Table A7). Each NDL program has 3 to 7 potential comparison programs, depending on the directorate and implementation year.

Table A7. Number of NDL and comparison programs, by directorate

Directorate	NDL implementation year(s)	Number of NDL programs	Number of comparison programs[a]
Geosciences (GEO)	2017 and 2018	2	6–7
Engineering (ENG)	2019	5	5
Biological Sciences (BIO)	2018 and 2019	5	3–4

Source: NSF databases.

[a]The number of comparison programs varies slightly by implementation year because of the requirement that comparison programs have six years of pre-implementation data and three years of post-implementation data.

Tables A8 through A10 show the NDL and comparison programs in each directorate (GEO, ENG, and BIO). As discussed in the memo, each comparison program receives a different weight depending on the NDL program with which it is being compared and the particular outcome of interest.

Table A8. GEO: NDL and comparison programs

Program number	Program title	NDL year	Number of proposals	Average project requested amount	Number of collaborative proposals per project	Number of reviewers	Proposal quality: Exceptional	Proposal quality: Fair	Proposal quality: Underperforming
NDL programs									
NSF 21-526	Arctic Research Opportunities	2017	397	$864,729	0.80	15.7	34%	57%	3%
NSF 21-567	Antarctic Research	2018	373	$799,019	0.90	18.3	46%	49%	1%
Comparison programs									
NSF 16-572	Cooperative Studies of the Earth's Deep Interior	n.a.	16	$466,552	0.63	13.4	60%	29%	0%
NSF 17-582	Paleo Perspectives on Climate Change	n.a.	184	$517,834	0.76	23.4	42%	50%	3%
NSF 20-579	Dynamics of Integrated Socio-Environmental Systems	n.a.	176	$1,317,097	0.01	25.7	19%	62%	9%
PD 98-1610	Physical Oceanography	n.a.	231	$869,953	0.56	19.8	37%	59%	1%
PD 98-1650	Biological Oceanography	n.a.	364	$828,551	0.70	22.4	34%	56%	1%
PD 98-1670	Chemical Oceanography	n.a.	211	$717,143	0.62	19.1	50%	41%	1%
PD 98-1680	Ocean Technology and Interdisciplinary Coordination	n.a.	38	$810,451	0.40	10.1	41%	50%	1%

Note: This table includes average program characteristics for the three years prior to NDL implementation for the two NDL programs and prior to 2018 for the seven comparison programs.

n.a. = not applicable.

Table A9. ENG: NDL and comparison programs

Program number	Program title	NDL year	Number of proposals	Average project requested amount	Number of collaborative proposals per project	Number of reviewers	Proposal quality: Exceptional	Proposal quality: Fair	Proposal quality: Underperforming
NDL programs									
PD 18-1517	Electronics, Photonics, and Magnetic Devices	2019	464	$411,737	0.13	12.0	19%	71%	4%
PD 18-7564	Communications, Circuits, and Sensing-Systems	2019	295	$430,247	0.20	12.4	13%	76%	5%
PD 18-7607	Energy, Power, Control, and Networks	2019	283	$438,393	0.21	10.5	11%	64%	18%
PD 20-5342	Disability and Rehabilitation Engineering	2019	70	$342,220	0.13	10.9	18%	54%	15%
PD 20-7909	Biosensing	2019	133	$358,156	0.12	13.0	11%	69%	8%
Comparison programs									
NSF 19-506	Partnerships for Innovation	n.a.	311	$481,176	0.00	8.0	9%	79%	10%
NSF 20-558	PFE: Research Initiation in Engineering Formation	n.a.	49	$181,953	0.12	9.0	13%	75%	10%
NSF 20-570	Industry-University Cooperative Research Centers Program	n.a.	208	$285,335	0.00	7.6	18%	35%	1%
NSF 21-539	Spectrum and Wireless Innovation Enabled by Future Technologies	n.a.	144	$906,765	0.61	10.4	11%	79%	8%
NSF 21-615	Emerging Frontiers in Research and Innovation	n.a.	142	$1,978,084	0.00	10.9	30%	64%	5%

Note: This table includes average program characteristics for the three years prior to NDL implementation (2019).

n.a. = not applicable.

Table A10. BIO: NDL and comparison programs

Program number	Program title	NDL year	Number of proposals	Average project requested amount	Number of collaborative proposals per project	Number of reviewers	Proposal quality: Exceptional	Proposal quality: Fair	Proposal quality: Underperforming
NDL programs									
NSF 21-504	Division of Environmental Biology (core programs)	2018	2,058	$158,336	0.13	24.7	37%	51%	6%
NSF 21-506	Division of Integrative Organismal Systems Core Programs	2018	2,245	$184,612	0.07	18.9	35%	56%	5%
NSF 21-544	Long Term Research in Environmental Biology	2018	89	$92,440	0.09	23.6	33%	49%	5%
NSF 21-501	Infrastructure Capacity for Biological Research	2019	208	$568,114	0.18	16.5	30%	61%	5%
NSF 21-509	Division of Molecular and Cellular Biosciences: Investigator-initiated research projects	2019	911	$865,738	0.14	16.6	32%	57%	5%
Comparison programs									
NSF 15-576	Advancing Digitization of Biodiversity Collections	n.a.	69	$1,141,407	2. 48	15.0	43%	52%	3%
NSF 21-545	Dimensions of Biodiversity FY2021	n.a.	209	$1,737,984	1.11	34.4	29%	64%	2%
NSF 21-609	Ecology and Evolution of Infectious Diseases	n.a.	64	$2,167,897	0.00	18.5	17%	67%	7%
NSF 20-579	Dynamics of Integrated Socio-Environmental Systems	n.a.	192	$1,332,711	0.00	23.2	18%	62%	10%

Note: This table includes average program characteristics for the three years prior to NDL implementation for the five NDL programs and prior to 2019 for the four comparison programs.

n.a. = not applicable.

B. Synthetic control analysis process

To construct the synthetic control groups, we used the augmented synthetic control R package (augsynth) using ridge regression (v0.2.0; Ben-Michael, Feller, and Rothstein 2021).[29] Augmented synthetic controls can improve pre-treatment fit by allowing for a broader range of weights than traditional synthetic controls. However, not imposing any restrictions on weights can introduce extrapolation bias. Augmented synthetic controls with ridge penalizes distance from synthetic control weights, thus allowing for a broader set of weights while minimizing extrapolation bias. Augmented synthetic controls with ridge regression has been shown to reduce bias and root mean squared error relative to traditional synthetic controls (Ben-Michael et al. 2021).

The ridge-augmented synthetic control estimator developed by Ben-Michael and colleagues (2021) is the following:

$$\hat{Y}_{1T}^{aug}(0) = \sum_{W_i=0} \hat{\gamma}_i^{aug} Y_{iT}$$

where:

$$\hat{\gamma}_i^{aug} = \hat{\gamma}_i^{sc} + (\boldsymbol{X}_1 - \boldsymbol{X}_0'\hat{\boldsymbol{\gamma}}^{sc})'\left(\boldsymbol{X}_0'\boldsymbol{X}_0 + \lambda^{ridge} I_{T0}\right)^{-1}\boldsymbol{X}_i.$$

Where:

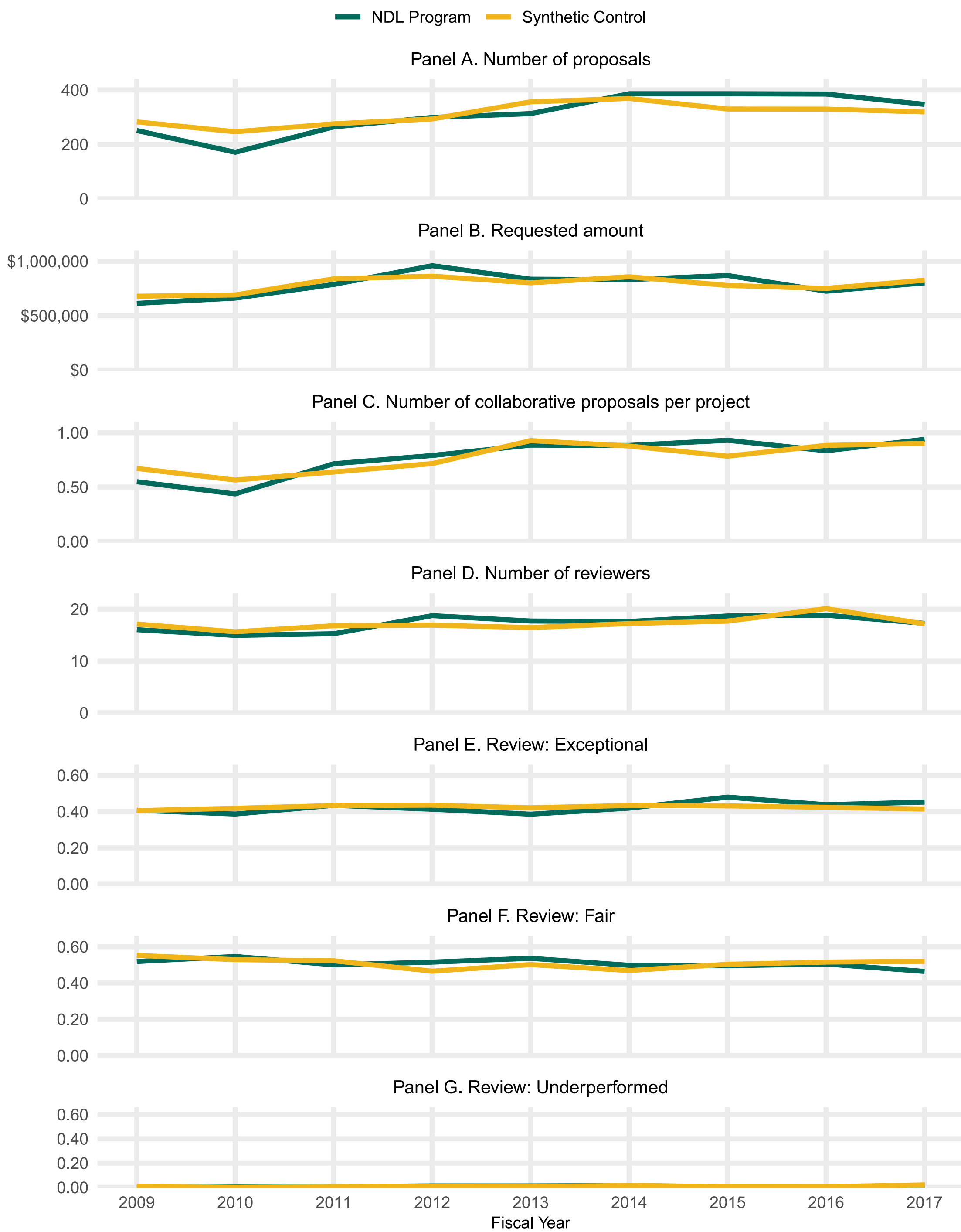
- $\hat{Y}_{1T}^{aug}(0)$ represents the augmented synthetic control estimator
- $W_i = 0$ represents units that do not implement NDL
- $\hat{\gamma}_i^{aug}$ represents the weight for the augmented synthetic control estimator for unit *i*
- Y_{iT} represents post-treatment outcomes for unit *i*
- $\hat{\gamma}_i^{sc}$ represents the weight for the traditional synthetic control estimator for unit *i*
- $\boldsymbol{X}_1$ represents the matrix of NDL pre-treatment outcomes
- $\boldsymbol{X}_0$ represents the matrix of non-NDL pre-treatment outcomes
- λ^{ridge} represents the penalty hyperparameter, which is determined through cross-validation
- I_{T0} represents the identity matrix

We assessed model fit in two primary ways: (1) by visually examining fit for pre-treatment years and (2) by examining fit metrics.[30] Figures A1–A12 show fit for pre-treatment years for all 12 programs. For a strong match, we examine whether the synthetic control follows the NDL program closely in the pre-treatment years. For example, in Figure A1, we look for a close fit from FY 2009 to FY 2017 (the year before implementation began) for all outcomes.

[29] Procedures were conducted using R 4.0.2 and the tidyverse (v1.3.0; Wickham et al. 2019).
[30] We show L^2 here, which is the metric used by Ben-Michael et al. (2021). Root mean square prediction error is another common way to assess model fit, and could be examined in future analyses (see, for example, Abadie 2015).

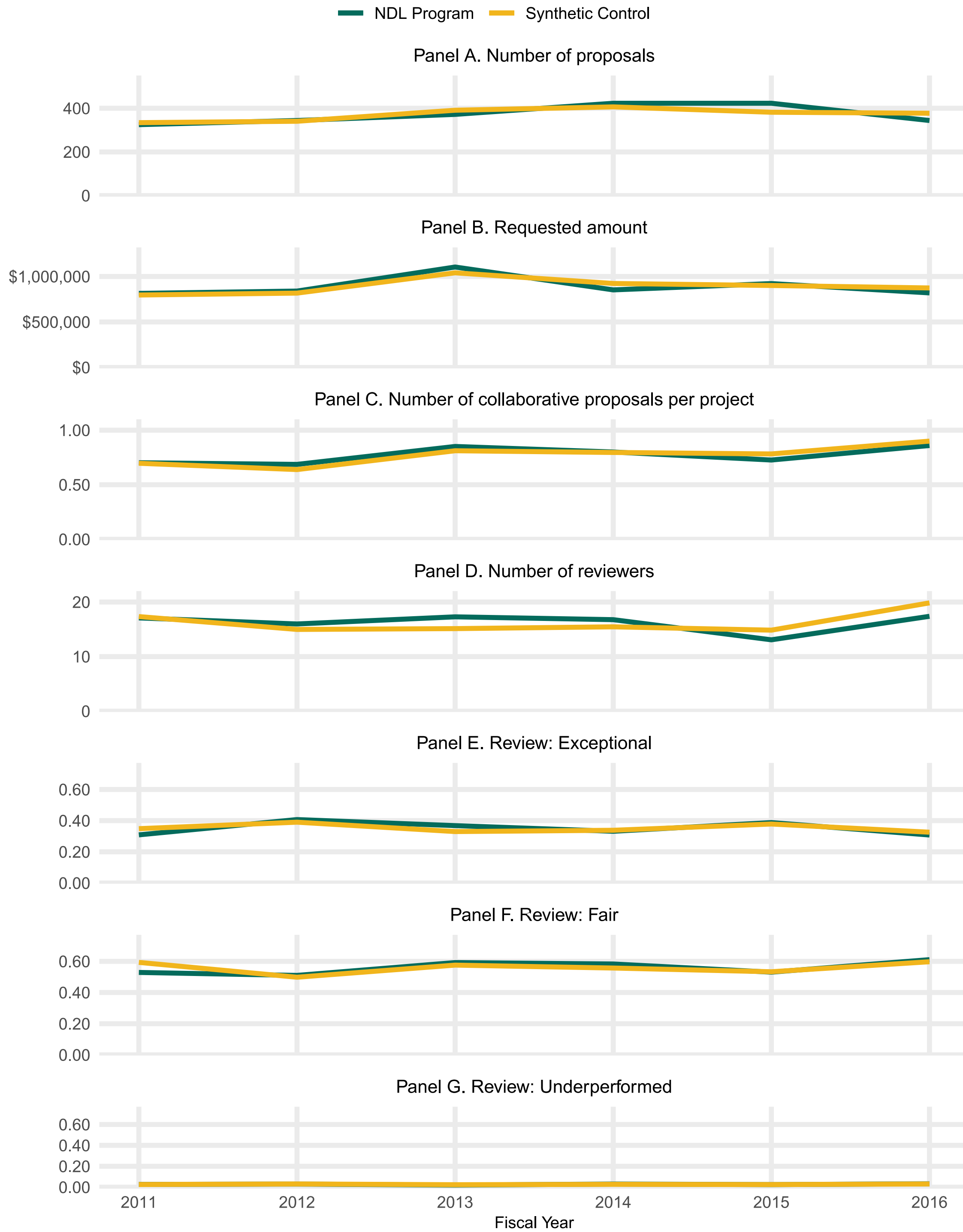
In Figures A1 through A5, there are relatively close matches for all pre-treatment years, indicating that the synthetic control approach was able to construct comparison groups that trend similarly to the NDL programs. In Figures A6 through A12, the pre-treatment trends diverge more between the NDL programs and synthetic control programs. For example, the pre-implementation trends for the Division of Integrative Organismal Systems Core NDL Program (NSF 21-506) did not vary substantially from year to year, but its estimated counterfactual exhibits some sharp increases and decreases in the pre-implementation years (see Figure A10). This is likely because there were only three to four potential comparison programs for BIO, making it difficult to estimate counterfactuals that closely follow outcome trends over the pre-implementation years. Indeed, BIO programs generally achieved the worst fit statistics as well (see Table A10 in the technical appendix).

Figure A1. NDL and synthetic control trends before NDL implementation for GEO: Antarctic Research Opportunities (NSF 21−567)



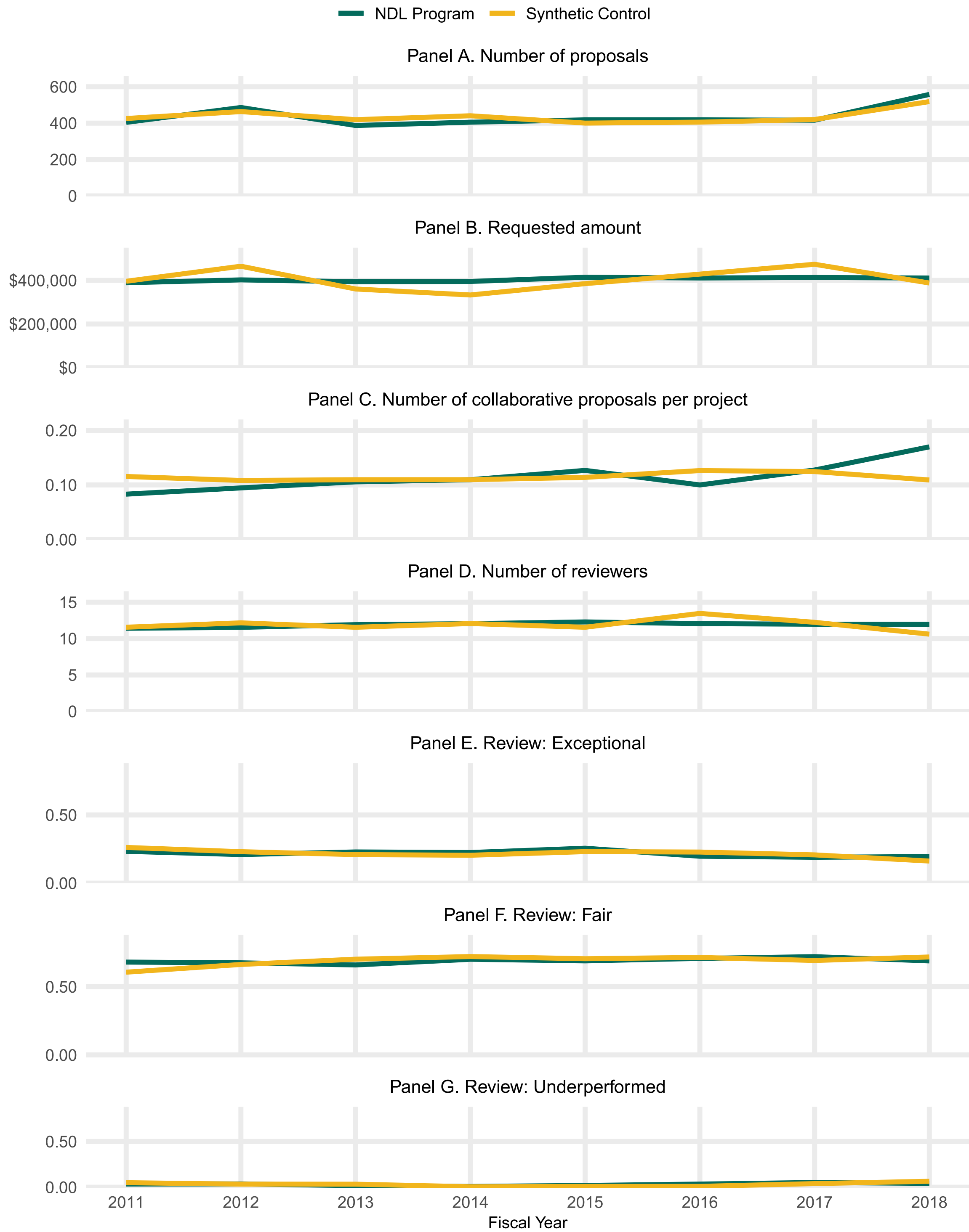
Note: The synthetic controls are weighted averages of the programs in GEO that have at least six years of data before NDL implementation, at least three years of data after NDL implementation, and did not implement NDL during this period. This includes seven comparison programs (see Table A8). The weights differ by outcome. There are 1,562 projects in the NDL program and 7,152 projects across the programs in the synthetic control prior to NDL implementation.

Figure A2. NDL and synthetic control trends before NDL implementation for GEO: Arctic Research Opportunities (NSF 21−526)



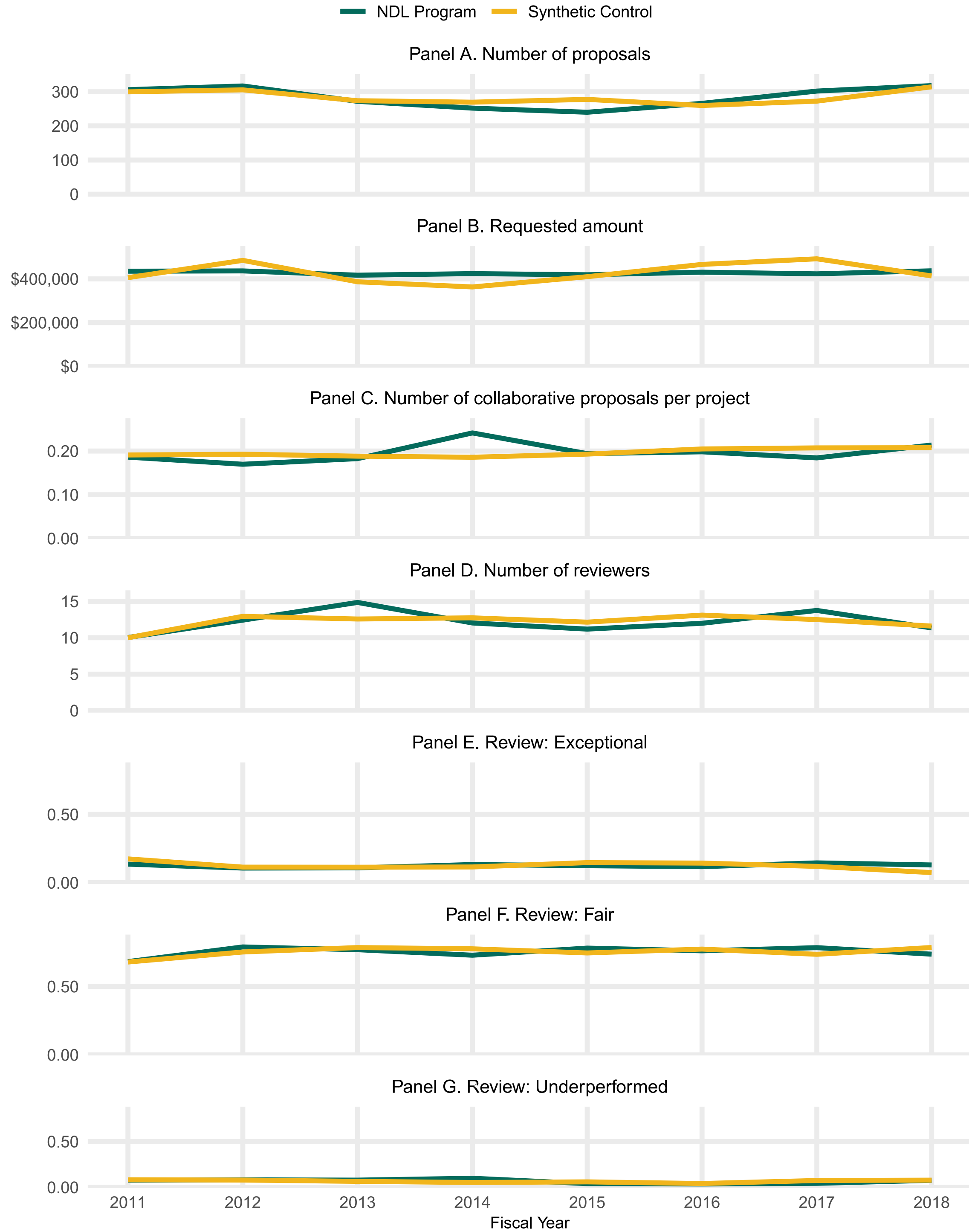
Note: The synthetic controls are weighted averages of the programs in GEO that have at least six years of data before NDL implementation, at least three years of data after NDL implementation, and did not implement NDL during this period. This includes six comparison programs (see Table A8). The weights differ by outcome. There are 1,261 projects in the NDL program and 4,030 projects across the programs in the synthetic control prior to NDL implementation.

Figure A3. NDL and synthetic control trends before NDL implementation for ENG: Electronics, Photonics and Magnetic Devices (PD 18-1517)



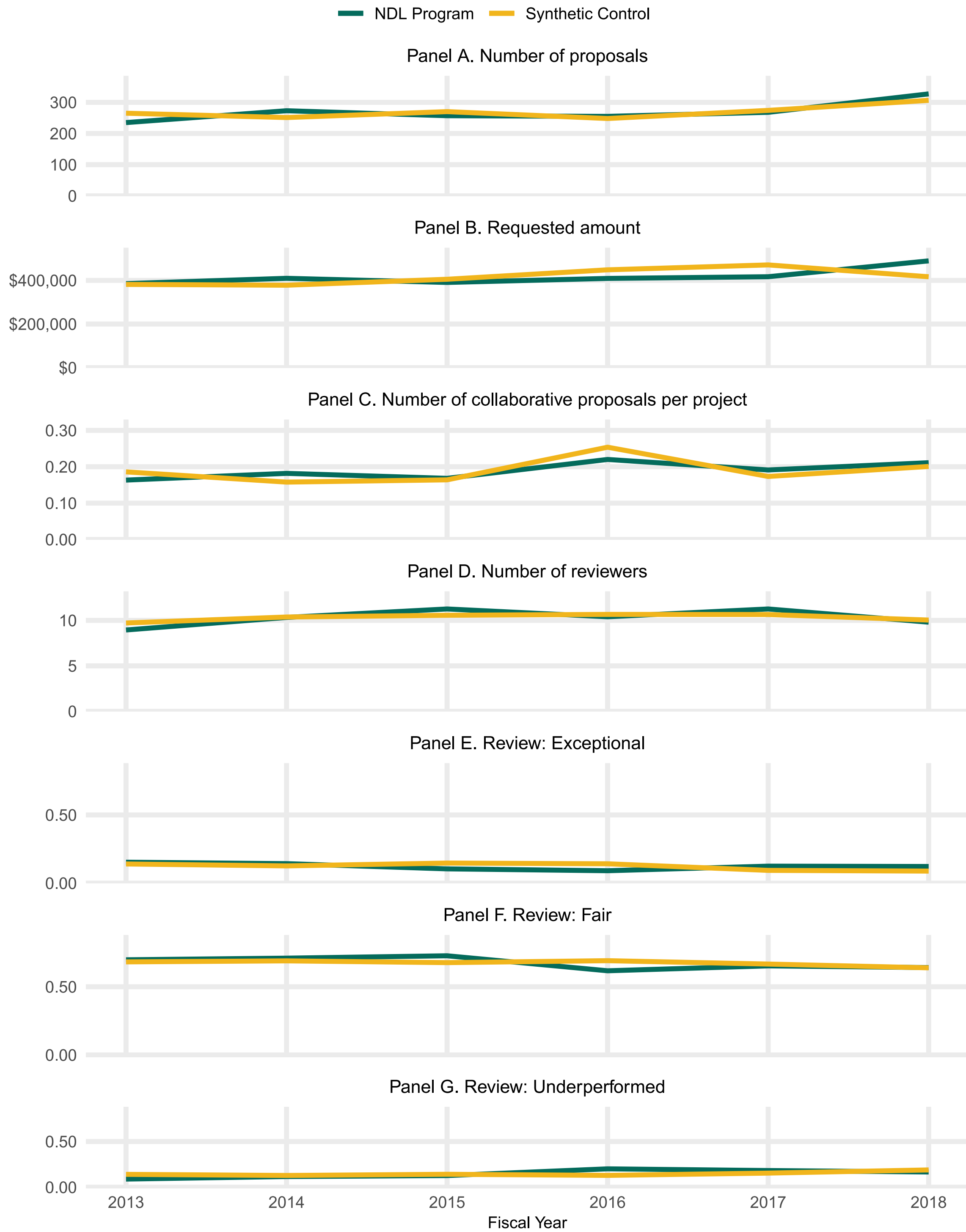
Note: The synthetic controls are weighted averages of the programs in ENG that have at least six years of data before NDL implementation, at least three years of data after NDL implementation, and did not implement NDL during this period. This includes five comparison programs (see Table A9). The weights differ by outcome. There are 3,129 projects in the NDL program and 5,371 projects across the programs in the synthetic control prior to NDL implementation.

Figure A4. NDL and synthetic control trends before NDL implementation for ENG: Communications, Circuits, and Sensing-Systems (PD 18-7564)



Note: The synthetic controls are weighted averages of the programs in ENG that have at least six years of data before NDL implementation, at least three years of data after NDL implementation, and did not implement NDL during this period. This includes five comparison programs (see Table A9). The weights differ by outcome. There are 1,902 projects in the NDL program and 5,371 projects across the programs in the synthetic control prior to NDL implementation.

Figure A5. NDL and synthetic control trends before NDL implementation for ENG: Energy, Power, Control, and Networks (PD 18-7607)

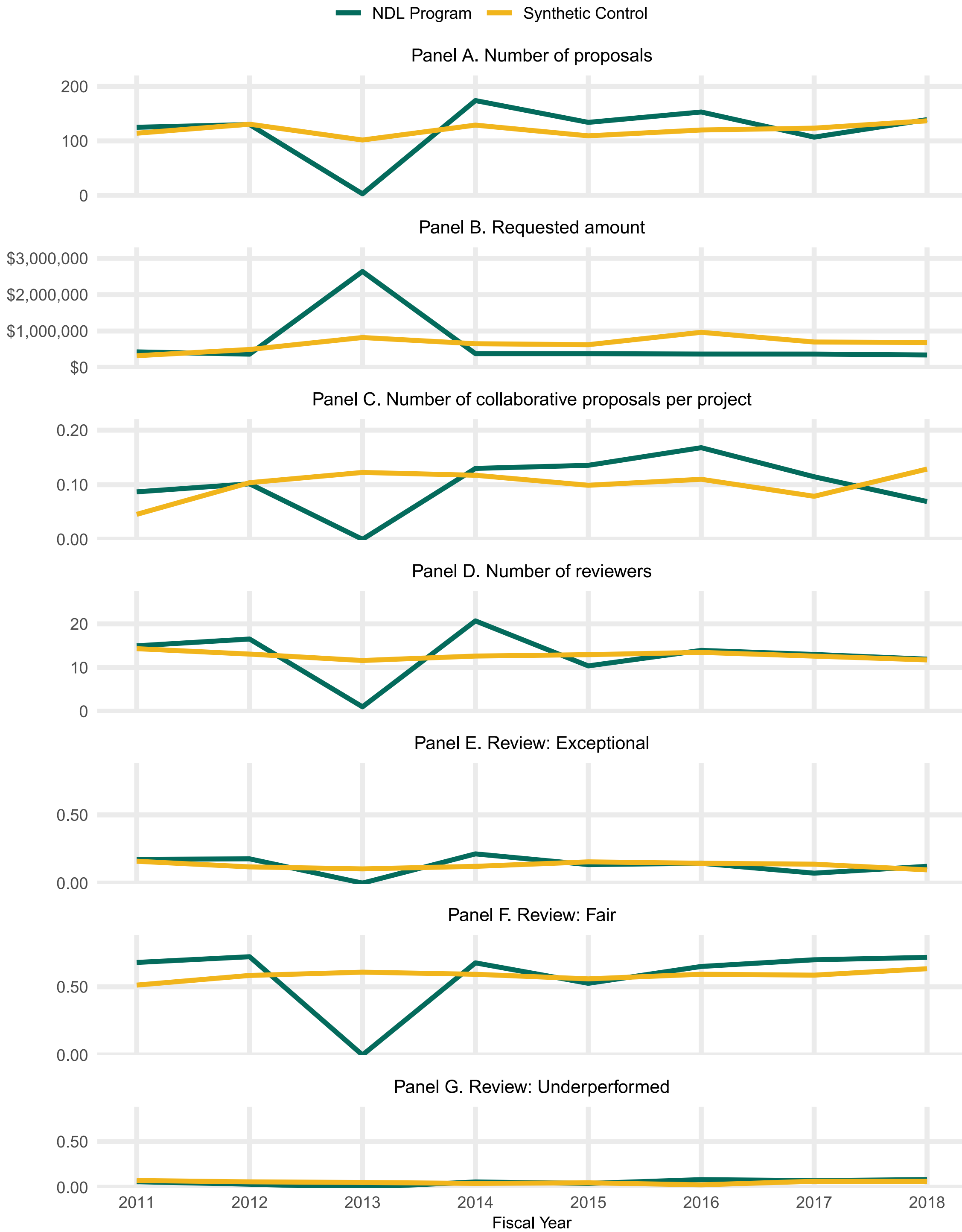


Note: The synthetic controls are weighted averages of the programs in ENG that have at least six years of data before NDL implementation, at least three years of data after NDL implementation, and did not implement NDL during this period. This includes five comparison programs (see Table A9). The weights differ by outcome. There are 1,357 projects in the NDL program and 4,305 projects across the programs in the synthetic control prior to NDL implementation.

Figure A6. NDL and synthetic control trends before NDL implementation for ENG: Disability and Rehabilitation Engineering (PD 20-5342)

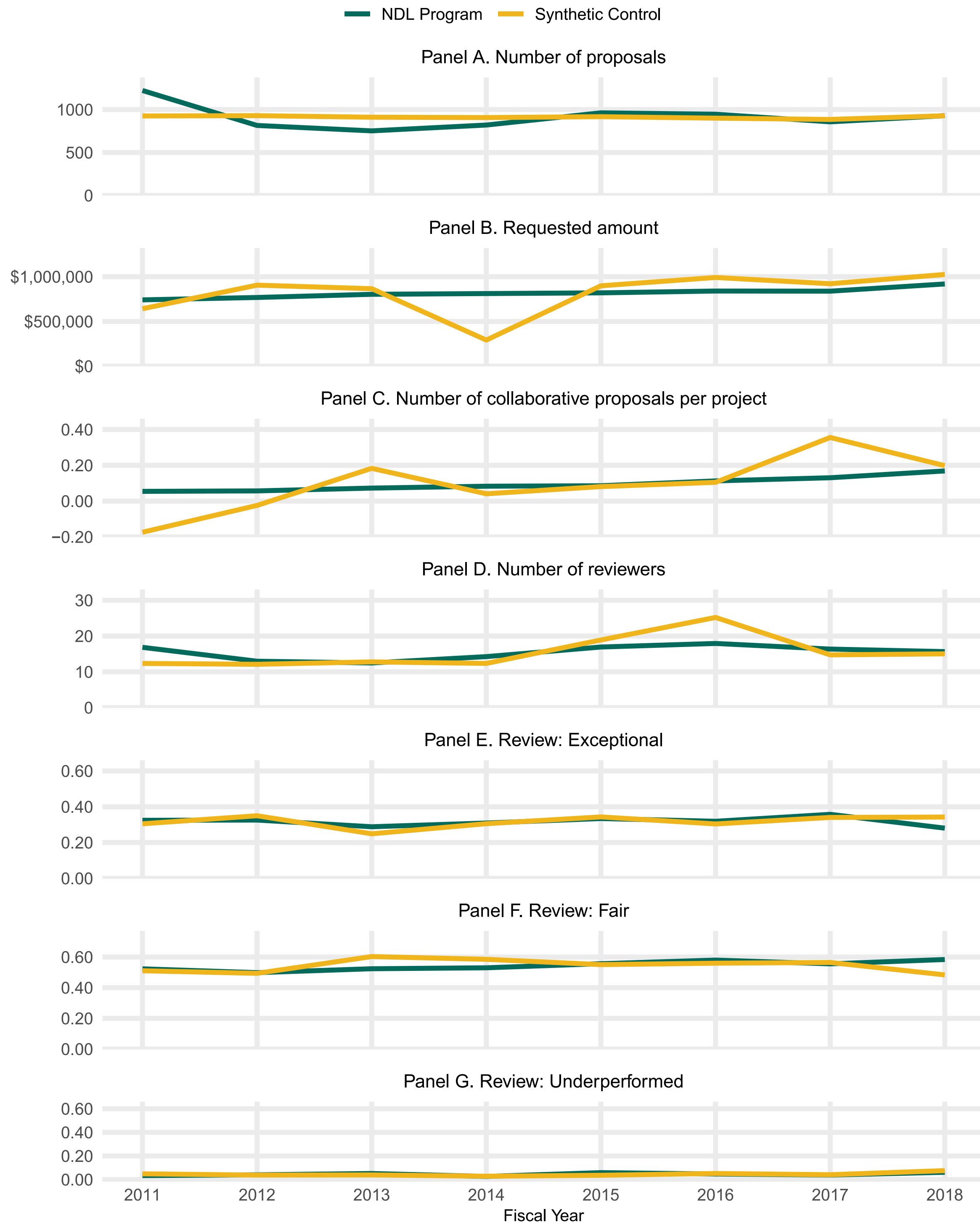
Note: The synthetic controls are weighted averages of the programs in ENG that have at least six years of data before NDL implementation, at least three years of data after NDL implementation, and did not implement NDL during this period. This includes five comparison programs (see Table A9). The weights differ by outcome. There are 445 projects in the NDL program and 5,371 projects across the programs in the synthetic control prior to NDL implementation.

Figure A7. NDL and synthetic control trends before NDL implementation for ENG: Biosensing (PD 20-7909)



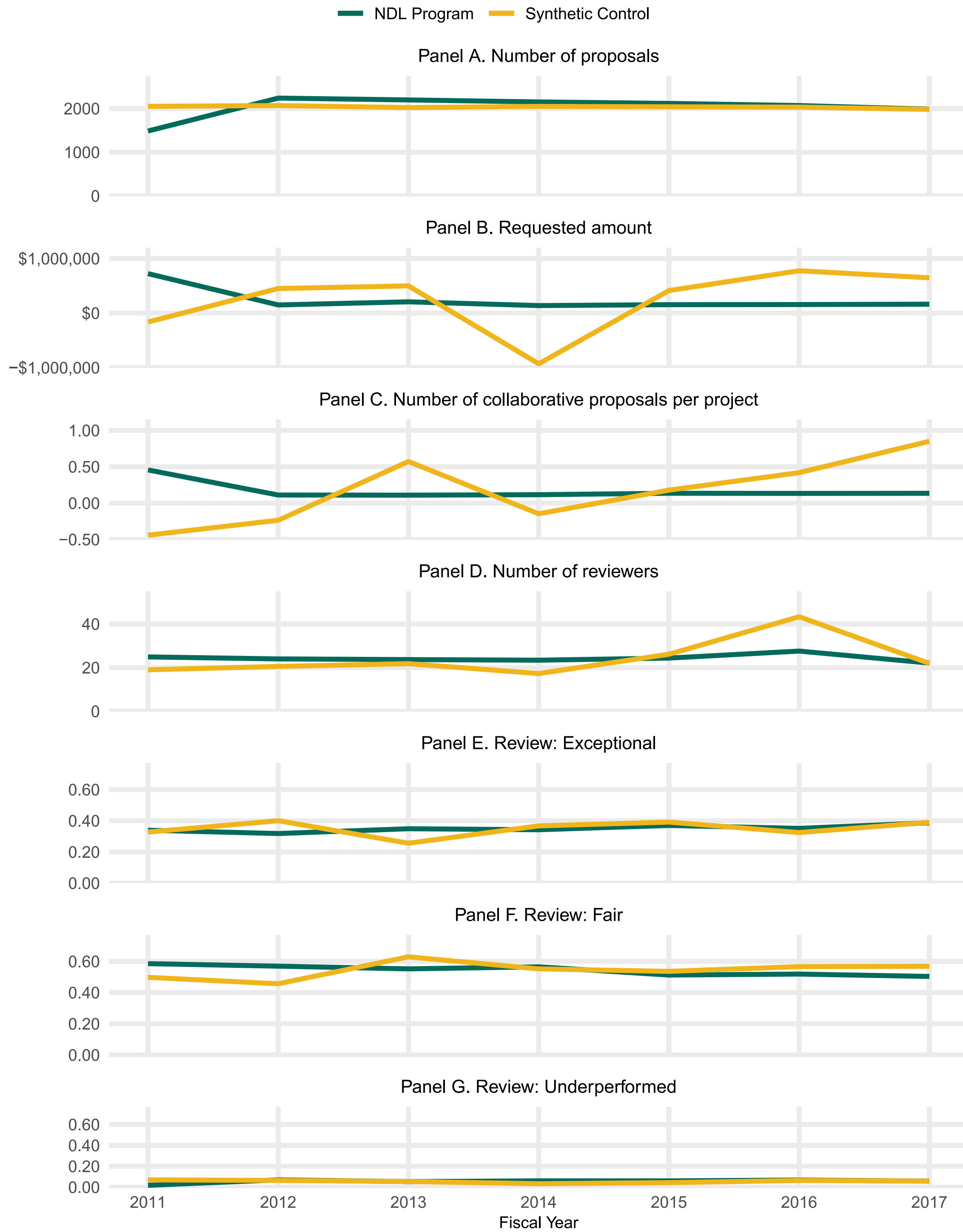
Note: The synthetic controls are weighted averages of the programs in ENG that have at least six years of data before NDL implementation, at least three years of data after NDL implementation, and did not implement NDL during this period. This includes five comparison programs (see Table A9). The weights differ by outcome. There are 865 projects in the NDL program and 5,371 projects across the programs in the synthetic control prior to NDL implementation.

Figure A8. NDL and synthetic control trends before NDL implementation for BIO: Division of Molecular and Cellular Biosciences Investigator-initiated research projects (NSF 21-509)



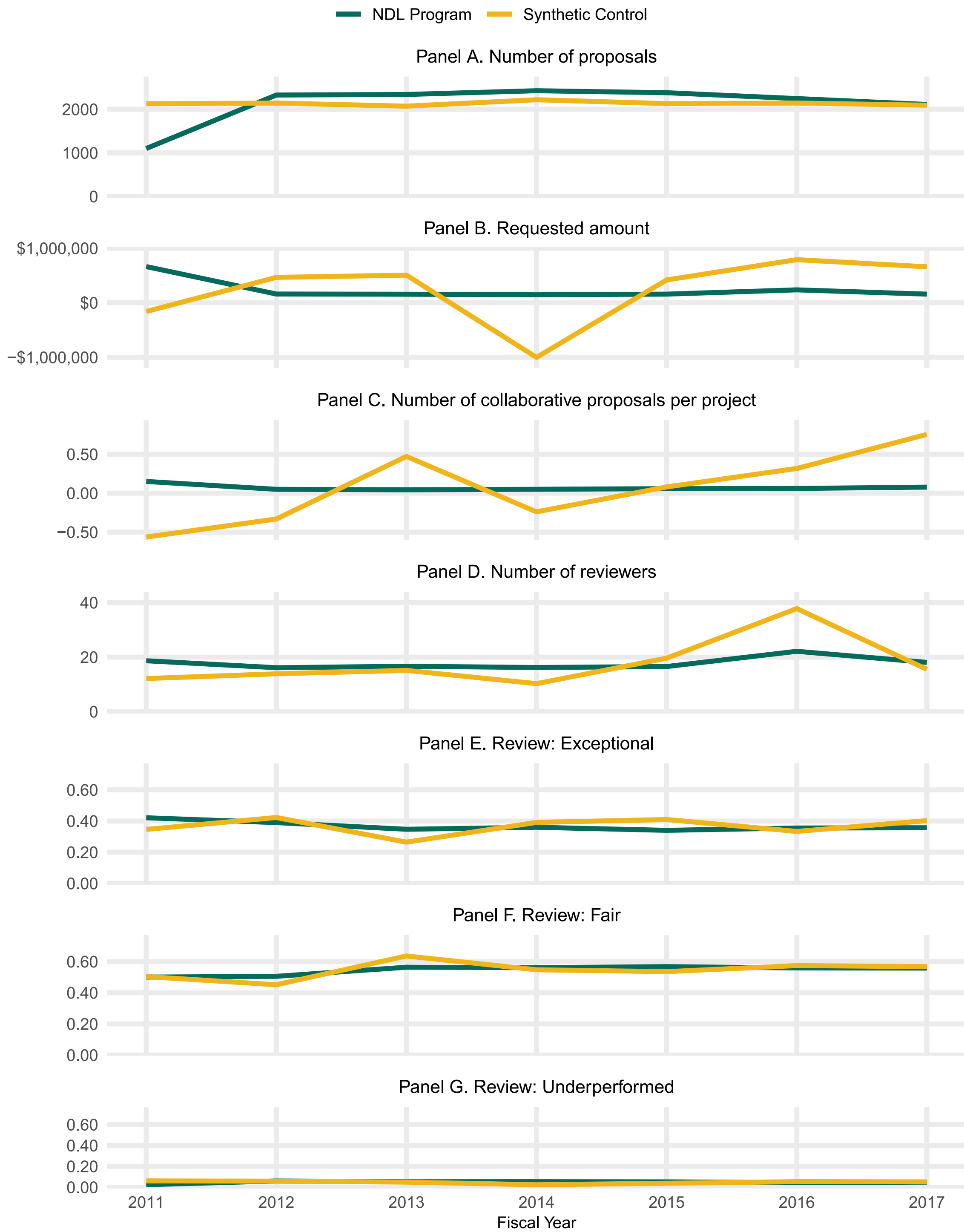
Note: The synthetic controls are weighted averages of the programs in BIO that have at least six years of data before NDL implementation, at least three years of data after NDL implementation, and did not implement NDL during this period. This includes four comparison programs (see Table A10). The weights differ by outcome. There are 6,677 projects in the NDL program and 3,124 projects across the programs in the synthetic control prior to NDL implementation.

Figure A9. NDL and synthetic control trends before NDL implementation for BIO: Division of Environmental Biology (core programs) (NSF 21-504)



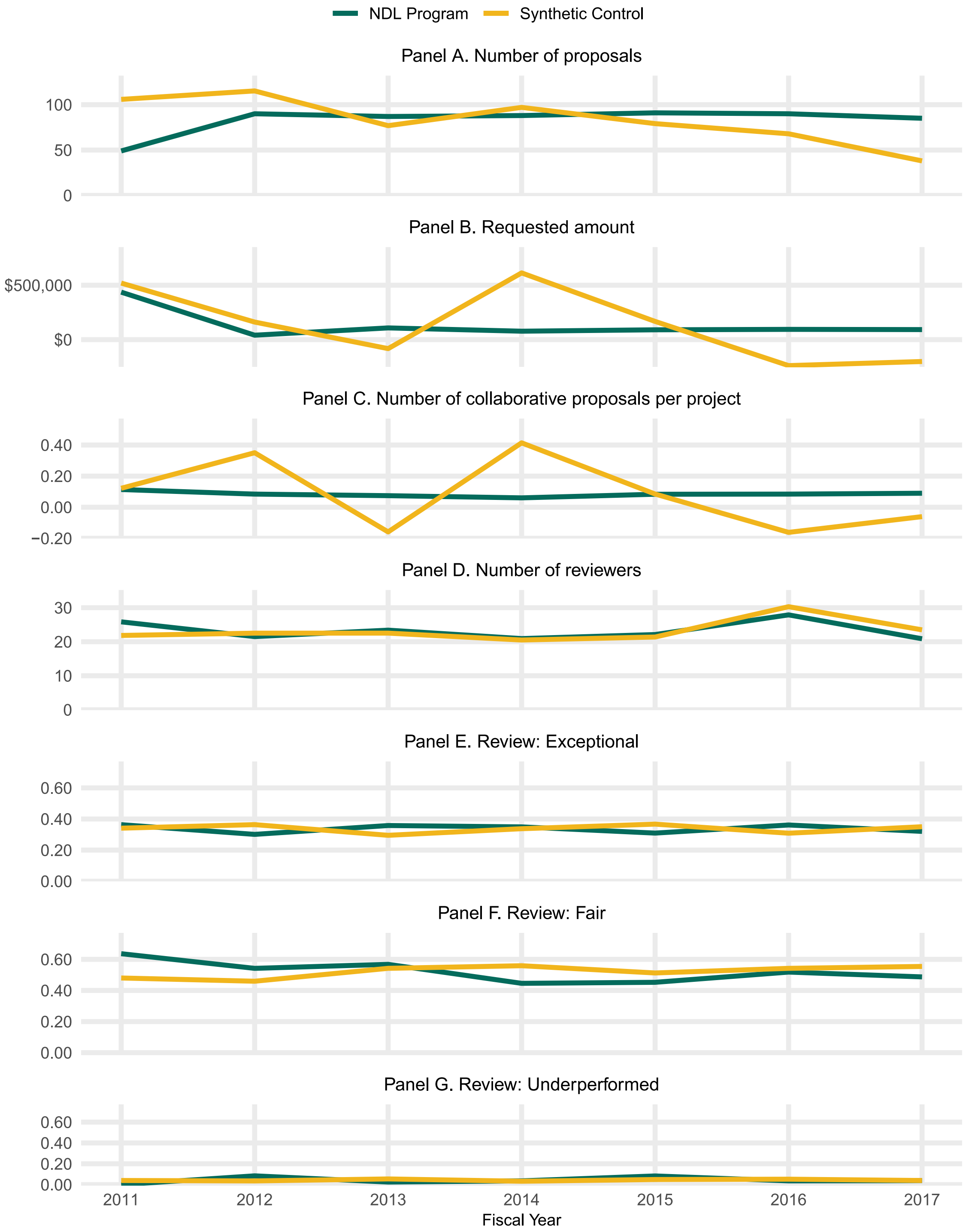
Note: The synthetic controls are weighted averages of the programs in BIO that have at least six years of data before NDL implementation, at least three years of data after NDL implementation, and did not implement NDL during this period. This includes three comparison programs (see Table A10). The weights differ by outcome. There are 12,389 projects in the NDL program and 1,605 projects across the programs in the synthetic control prior to NDL implementation.

Figure A10. NDL and synthetic control trends before NDL implementation for BIO: Division of Integrative Organismal Systems Core Programs (NSF 21-506)



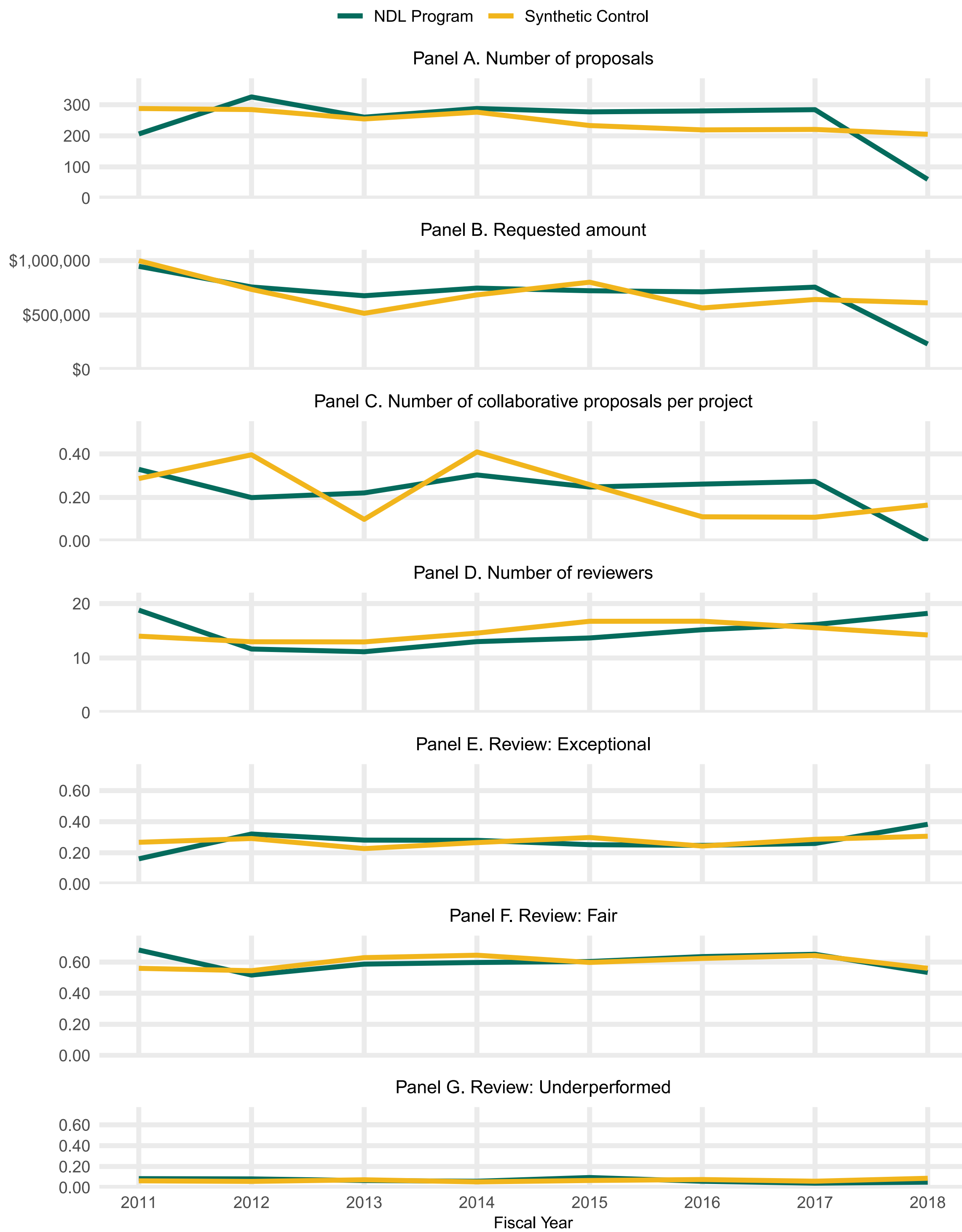
Note: The synthetic controls are weighted averages of the programs in BIO that have at least six years of data before NDL implementation, at least three years of data after NDL implementation, and did not implement NDL during this period. This includes three comparison programs (see Table A10). The weights differ by outcome. There are 14,046 projects in the NDL program and 1,605 projects across the programs in the synthetic control prior to NDL implementation.

Figure A11. NDL and synthetic control trends before NDL implementation for BIO: Long Term Research in Environmental Biology (NSF 21-544)



Note: The synthetic controls are weighted averages of the programs in BIO that have at least six years of data before NDL implementation, at least three years of data after NDL implementation, and did not implement NDL during this period. This includes three comparison programs (see Table A10). The weights differ by outcome. There are 536 projects in the NDL program and 1,605 projects across the programs in the synthetic control prior to NDL implementation.

Figure A12. NDL and synthetic control trends before NDL implementation for BIO: Infrastructure Capacity for Biological Research (NSF 21-501)



Note: The synthetic controls are weighted averages of the programs in BIO that have at least six years of data before NDL implementation, at least three years of data after NDL implementation, and did not implement NDL during this period. This includes four comparison programs (see Table A10). The weights differ by outcome. There are 1,587 projects in the NDL program and 3,124 projects across the programs in the synthetic control prior to NDL implementation.

Table A11 shows the scaled L^2 distance between the NDL program and its synthetic control. L^2 represents the difference between the pre-treatment outcomes for the NDL program compared to the synthetic control. Ben-Michael et al. (2021) define this as:

$$L^2 = \sqrt{\Sigma\left(\boldsymbol{X}_0\widehat{\boldsymbol{\gamma}}_{aug} - \boldsymbol{X}_1\right)^2}$$

where:

- $\boldsymbol{X}_1$ represents the matrix of NDL pre-treatment outcomes
- $\boldsymbol{X}_0$ represents the matrix of non-NDL pre-treatment outcomes, and
- $\widehat{\boldsymbol{\gamma}}^{aug}$ represents the weights for the augmented synthetic control estimator.

We present the scaled L^2, which is scaled by the imbalance using uniform weights. Values of 1 or greater indicate that there are not improvements over using uniform weights. Several of the BIO programs have relatively poor fits, particularly for the average project requested amount outcome (with L^2 values greater than 1).

Table A11. Quality of the synthetic control fit: scaled L^2 distance between the NDL program and its synthetic control, by outcome

Directorate	NDL program	Number of proposals	Average project requested amount	Proposal quality	Number of reviewers	Number of collaborative proposals per project
GEO	NSF 21-567	0.71	0.61	0.76	0.52	0.49
GEO	NSF 21-526	0.79	1.10	0.58	0.95	0.50
ENG	PD 18-1517	0.51	0.79	0.42	0.39	0.61
ENG	PD 18-7564	0.50	0.66	0.50	0.43	0.55
ENG	PD 18-7607	0.98	0.37	0.84	0.28	0.63
ENG	PD 20-5342	0.73	0.78	0.84	0.60	0.96
ENG	PD 20-7909	0.82	0.90	0.86	0.72	0.96
BIO	NSF 21-509	1.00	2.35	1.11	0.31	0.66
BIO	NSF 21-504	0.98	6.44	1.25	1.02	1.42
BIO	NSF 21-506	0.97	8.31	1.13	0.89	1.43
BIO	NSF 21-544	0.92	1.38	0.88	0.42	0.65
BIO	NSF 21-501	0.88	0.97	0.73	0.37	0.56

Source: augsynth model output.

C. Exploratory outcomes analysis

Table A12 shows the average change in outcomes after NDL implementation for the two chosen NDL programs (Antarctic Research and Artic Research Opportunities). After NDL was implemented, on average NDL programs received 123 to 128 fewer proposals, NDL projects requested $265,000 to $294,000 more funding, NDL proposals required 0.9 to 1.8 fewer reviewers, and proposals were less likely to be rated as fair (by 4 to 9 percentage points) or underperforming (by one percentage point). There were less consistent patterns for number of collaborative proposals per project and exceptional reviews.

Across the two NDL programs, there is conflicting evidence as to whether differences (for the NDL programs relative to the synthetic controls) may be increasing or decreasing after the first year. Differences increased in magnitude after the first year for a few of the outcomes (including number of proposals, requested amount, and underperformed proposals for NSF 21-526) and decreased in magnitude for all other outcomes. We did not conduct any statistical tests for whether differences may differ over time.

Table A12. Average change in outcomes after NDL implementation for Antarctic Research (NSF 21−567) and Arctic Research Opportunities (NSF 21-567)

	NSF 21-526		NSF 21-567	
	Average change	Change in first year	Average change	Change in first year
Number of proposals	-122.59	-101.14	-127.64	-202.70
Project requested amount	$294,291	$61,519	$264,926	$684,589
Number of collaborative proposals per project	0.062	-0.152	0.002	0.078
Number of reviewers	-1.83	-4.46	-0.85	-1.58
Review: exceptional	0.031	0.085	-0.010	0.075
Review: fair	-0.088	-0.207	-0.037	-0.163
Review: underperformed	-0.014	-0.008	-0.006	-0.009

Source: augsynth model output.

Note: The average change represents the average change in outcomes over all years after NDL implementation, and the year 1 change represents the change in outcomes in the first year after NDL implementation. Given that this is an exploratory analysis, standard errors are not presented.

V. References

Abadie, Alberto, Diamond, Alexis, and Hainmueller, Jens. 2015. “Comparative Politics and the Synthetic Control Method.” *American Journal of Political Science* 59(2):495-510.

Ben-Michael, Eli, Avi Feller, and Jesse Rothstein. 2021. “The Augmented Synthetic Control Method.” *Journal of the American Statistical Association* 116(536):1789-1803.

National Science Foundation. 2021. “Querying Proposal and Award Data Resources and Data Sources.” Prepared for New Employee Roundtable, October 2021.

Pankow, Neha. 2020. *Removal of Deadlines for Programs across NSF: A Template for Case Studies*. Washington, DC: National Science Foundation.

Patino, Lina and Diana Hernandez Garcia. 2020. “Removing Proposal Submission Deadlines in Earth Sciences.” Presented to the National Science Foundation, December 2, 2020.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo, and Hiroaki Yutani. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4(43):1686. doi:10.21105/joss.01686.

Yuan, Grace, Gabriela Niño de Guzmán, Yuen Lau, Steven Zehnder, Richard Nash, Jesus Alvelo Maurosa, and Miriam Scheiber. 2020. "ENG and No Deadlines." Presented to the National Science Foundation, December 3, 2020.