Supporting Statement B

National Institutes of Health
Evaluation of the Enhancing Diversity of the NIH-funded Workforce Program
OMB Control Number 0925-0747

January 14, 2021

Alison Gammie, Ph.D.
Director, Division of Training, Workforce Development, and Diversity
Project Lead, Diversity Program Consortium
National Institute of General Medical Sciences
45 Center Drive, 2AS43H
Bethesda, MD 20892
Phone: (301) 594-2662
Email: alison.gammie@nih.gov

Contents

List of Attachments

B. Collection of Information Employing Statistical Methods

This OMB *reinstate with revisions* application seeks approval for 3-year clearance to continue an evaluation of the NIH Common Fund's Enhancing Diversity of the NIH-funded Workforce Program (also referred to as the Diversity Program Consortium) - a national consortium comprised of three integrated initiatives: (1) Building Infrastructure Leading to Diversity (BUILD), (2) National Research Mentoring Network (NRMN), and (3) Coordination and Evaluation Center (CEC). We are requesting OMB clearance for the data collection that is required for the CEC to evaluate the overall impact and effectiveness of the BUILD and NRMN initiatives as required by the NIH Common Fund award. The evaluation will assess agreed-upon consortium-wide hallmarks of success at the student/mentee, faculty/mentor, and institutional level (see Attachment 6, 7, and 8).

B.1 Respondent Universe and Sampling Methods

The longitudinal evaluation uses mixed methods to assess implementation and outcomes of the Diversity Program Consortium. Data collection includes annual surveys, institutional and program data collection, along with site visits and case studies finding as detailed in **Section A** of this application.

B.1.1. Respondent Universe

The respondent universe varies depending on the outcome of interest (e.g., BUILD students, BUILD faculty, participants in various NRMN activities).

BUILD Students: The CEC surveys students identified by each of the 10 BUILD programs as being a participant in their program, starting when they enter the program (usually as either a freshman or as a transfer student). We also sample additional students from each BUILD institution to achieve the target 500 students annually (this number is based on the power calculations presented in Section B.1.2 below). Entering freshmen and transfer students will be sampled non-proportionally with greater weight (80:20) given to those with a declared biomedical major (the focus of the BUILD initiative). In addition, we will over-sample African American and Hispanic/Native American subgroups considering their lower response rates as seen in annual UCLA Higher Education Research Institute (HERI) surveys (see Section A.2 for details on the HERI survey) . For African Americans, the oversampling rate will be 160%; for the Hispanic/Native American group it will be 130% relative to Whites/Others. This oversampling is to ensure that in each year, the resulting sample of students at each BUILD institution reflects the demographic characteristics of entering students at those institutions. BUILD institutions typically have populations with a high proportion of individuals from underrepresented groups, e.g., racial/ethnic minorities, low socioeconomic status (see Table in A.8.2. for list of awardee institutions). Students selected for inclusion in the evaluation data collection are asked to complete the "The Cooperative Institutional Research Program" (CIRP) HERI Freshman Survey (Attachment 12), the Student Annual Follow-Up Survey (Attachment 13), and College Senior Survey (Attachment 14). For each of the 10 BUILD institutions, the

CEC has identified a comparable non-BUILD institution that has also collected HERI data. The matching of BUILD/non-BUILD institutions is based on student demographics and institutional characteristics. Secondary data from HERI surveys at those non-BUILD institutions will be used for comparison with outcomes at BUILD institutions.

BUILD Faculty (Attachment 15): We survey 500 faculty from all BUILD institutions combined (~50/institution) and 500 faculty from non-BUILD institutions (~50/institution). Faculty are sampled for baseline data only once per 3-year cycle as there is expected to be little change in faculty over that time based on faculty retention and turnover rates. At BUILD institutions, faculty are sampled such that all faculty participating in the BUILD program activities are included (unless there are more than 25, in which case a random sample will be drawn so that the faculty who have participated in BUILD do not represent more than 50% of the total sample of 50). In addition, a random sample of other faculty in biomedical disciplines will be drawn to complete the total sample of 50. Faculty at non-BUILD institutions will be randomly sampled from existing secondary data from HERI Faculty Surveys administered independent of this initiative by various US academic institutions.

NRMN: Phase I NRMN student participants (undergraduate through post-doctoral) and faculty/professionals at all levels are invited to complete the NRMN Annual Follow-up Survey (Attachment 17). The respondent universe relies on the participants' roles in the NRMN intervention, e.g., mentor training, grant writing skills workshops, and linking and matching activities. Provided the response rates are robust, the samples will reflect the demographic characteristics of those participating in the various NRMN activities.

B.2 Procedures for the Collection of Information

A mapping document linking the data sources to the Hallmarks of Success are provided in Attachment 7.

B.2.1. Power Analysis and Estimation Procedure

At the conclusion of the longitudinal data collection, the analysis will compare BUILD exposed groups (students, faculty, institution) to non-BUILD exposed comparison groups, or will compare NRMN participants exposed to given activities (grant-writing workshop, mentor training) to NRMN participants who were not exposed to that activity (entered through the web portal). For the various key outcomes (defined in the Hallmarks; see Attachment 6), generalized mixed linear models will be used to test the hypothesis that the BUILD or NRMN interventions results in better outcomes for those participating in BUILD or NRMN activities. Models will test for the significance of an interaction term reflecting the "difference of differences" of the senior minus freshman scores for BUILD students versus similar students in non-BUILD schools. A statistically significant term in a positive direction would indicate success of BUILD, after adjusting for the covariates and clustering within institutions.

The power calculations are based on testing the following key hypothesis:

Students (BUILD)

H1*: The "****Biomedical Career Interest Scale score****" will improve over time more for students in the BUILD program than for student in the non-BUILD institution.*

For this hypothesis, the outcome variable is the *Biomedical Career Interest Scale* score collected in the freshman and senior years. This scale will be computed from the combined answers to the following questions: Do you intend to pursue a science-related research career? How many months since entering college (including summer) did you work on a professor's research project? Have you participated in an undergraduate research program? How often did a professor provide you with an opportunity to conduct research? How often have you met with an advisor/counselor about your career plans? Since many possible covariates could be used to predict any outcome, a two-step method will be used to derive the final model. First, we perform a univariate analysis of each covariate as it relates to the outcome and select those that are associated with it at the p=0.20 significance level or less. We then regress the outcome on these selected covariates using a specific best-subset regression approach, from which a generalized mixed linear model relating the outcome to the selected covariates is derived. In the model, we would include as covariates all those selected above as well as: X1 = (BUILD/non-BUILD school), X2 = Year (freshman/senior) and X3 = the product of X1 and X2 as their interaction. Thus, the interaction term is a "difference of differences" of the senior minus freshman score for BUILD versus non-BUILD schools. A statistically significant interaction term in a positive direction would indicate success of BUILD.

We test the null hypothesis that the average change in the biomedical career interest scale from freshman to senior years is the same in the BUILD and non-BUILD institutions versus the alternative that the BUILD students will show a higher increase in the average score. We define Y = change in score (senior – freshman) and define the effect size as: Effect size = [(Mean of Y | BUILD) - (Mean of Y | non-BUILD)]/standard deviation of Y. With significance level of 0.05, we select the sample size necessary to produce 0.8 power to identify a low effect size of 0.25. The necessary sample size is 253 students in each of the BUILD institutions and their comparable non-BUILD institutions. Allowing for annual non-response attrition of 20% over 3-4 years, we proposed to sample 500 students in each institution from each incoming cohort for follow-up. We will also sample 500 per year from each of these institutions in order to provide adequate sample sizes for subsequent cohorts to evaluate potential differences in shorter-term, interim outcomes (e.g., research self-efficacy, science identify, intent to pursue biomedical career) for these different cohorts as each is likely to be exposed to somewhat different BUILD offerings as those are modified over the funding period. By sampling 500/year from each institution, we also position the Diversity Program Consortium to examine longer-term outcomes of these different cohorts (with their likely somewhat different exposures). Furthermore, the annual samples will allow us to examine longitudinal trends in our outcomes.

This type of analysis will be used for any similar longitudinal change hypothesis with a continuous outcome and for any two groups, e.g., all BUILD students vs. non-BUILD students, or BUILD faculty vs. non-BUILD faculty, and so forth. In particular, we test the same

hypothesis comparing all BUILD students vs. all non-BUILD students. To account for the clustering effect (with institution being a cluster), we incorporate an intra-class correlation ranging from 0.001 to 0.04 (Adams, Gulliford, Ukoumunne, et al., 2004).[1] The following table shows the power for a two-sided alternative and an effect size of 0.25.

Intra-class correlation	0.001	0.005	0.01	0.02	0.03	0.04
Power	1.00	1.00	1.00	0.95	0.86	0.76

Thus, the power is at least 0.76 for all scenarios, and is near or equal to 1.0 for most scenarios. The power figures for higher effect sizes of 0.5 and 0.8 are all equal to 1.00.

Another hypothesis to be tested relates to entry into the biomedical fields. Specifically,

H2*: The proportion of students in BUILD institutions who graduate with an undergraduate degree in a biomedical science discipline will be higher than the proportion in otherwise similar non-BUILD institutions*.

The analysis to test this hypothesis will also be a generalized linear mixed model, but with a binary outcome. The selection of the model and the included covariates will follow the same lines as described for H1.

BUILD Faculty & NRMN faculty/mentors

For the analysis, we will use the *Mentoring Competency Assessment* scale pre- and post-training[2] to test the hypothesis:

H3: *The mean change of post – pre score is greater in the NRMN-trained group than in the control group*.

The analysis will follow the same approach described for hypotheses H1 & H2. Turning to faculty/mentor data analysis, NRMN had a total of ~8,000 individuals who either participated in activities or signed into the web portal with no intervention. To find the size of the control group, we use a conservative two-sided t-test with power = 0.80 and N1 = 500 NRMN-trained faculty. We ignore the clustering effect (intra-class correlation) since the faculty will largely be recruited from different universities, with at most two faculty members from any one institution. Allowing for 20% attrition per year over a 3-year follow-up, we use N1 = 320 to calculate the needed number of controls. An effect size of 0.40 was derived from data provided in the reference below from a similar study.[3] However, since we will be following this group after training, we use an effect size of 0.25 to account for the longer time effect. The required sample size is N2 = 208

[1] Adams G, Gulliford MC, Ukoumunne OC, et al. Patterns of intra-cluster correlation from primary care research to inform study design and analysis. J Clin Epidemiol, 2004; 57: 785–94.

[2] Pfund C, House S, et al. Training Mentors of Clinical and Translational Research Scholars: A Randomized Controlled Trial, Acad Med., 2014 May; 89(5): 774–782 doi:10.1097/ACM.0000000000000218.

controls. Allowing for 20% attrition per year over 3-4- years, we need to recruit at least 260 controls. To be cautious, we will recruit 500 controls as this also positions the Diversity Program Consortium to examine longer-term outcomes of these groups should additional funding be available to support follow-up.

For BUILD institutions' faculty, we will follow the same approach as for NRMN faculty. Therefore, we also need to survey 500 BUILD institution faculty in order to ensure that we have a minimum of 300 faculty at the 3-year follow-up (allowing for 20% attrition annually). These faculty will be sampled from among institutional faculty in biomedical research fields, taking 100% of those participating in BUILD activities (unless there are more than 25 in which case a random sample of 25 will be drawn) and random sampling the balance needed for the target sample size from amongst the rest of the bioscience faculty. Similar to BUILD students, the control faculty will be sampled from biomedical research faculty at comparable institutions without BUILD programs. Their number is also 500. As for students from non-BUILD institutions, data for faculty at non-BUILD institutions will come from secondary data available from HERI surveys independently administered at these institutions. The outcomes for these analyses include mentoring efficacy (similar to NRMN) as well as the research productivity (e.g., numbers of publications, grant submissions). Thus, some outcomes will be scale measures (continuous) and others are counts to be analyzed with Poisson-type methodology. All will use the generalized mixed model methods described earlier. Subgroup analyses will compare BUILD faculty to others in the same institutions.

B.2.1. Data Collection Procedures

Surveys. The student and faculty surveys are administered according to the schedule outlined in Table A.16. The primary modality for all survey administration is online. However, understanding that individual respondents may have different preferences or that institutional factors may facilitate different methods (such as a group-administration during student orientation or other group activity), scannable paper surveys will also be available for use as needed. All surveys are formatted to be completed as computer-assisted interviews (conducted by CEC interviewers) for respondents who prefer this modality. Regardless of modality, surveys are designed with skip patterns so that respondents are presented only with questions that are relevant for them. Consent procedures are implemented as indicated by the Institutional Review Board at the University of California, Los Angeles. For most online surveys, this is in the form of a screen after the introductory information that will indicate how data will be handled and the confidentiality of the responses, with contact information for CEC staff should respondents have questions or concerns prior to beginning the survey. All surveys will provide introductory information about the purpose of the survey and the expected time for completion. Invitations to participate in online surveys are provided through email, with an individualized link or access code provided for the respondent. Non-respondents will be prompted with four follow up emails every 5-7 days.

[3] Pfund C, House S, et al. Training Mentors of Clinical and Translational Research Scholars: A Randomized Controlled Trial, Acad Med., 2014 May; 89(5): 774–782 doi:10.1097/ACM.0000000000000218.

Site Visits & Case Studies. The CEC periodically conducts visits to awardee sites to gather qualitative data (see Attachment 18 for protocols). The data are coded, curated, and securely stored under the management and oversight of the CEC.

BUILD Participant Data. BUILD awardees submit participant rosters on an ongoing basis through the CEC Tracker, a tool developed by the CEC and utilized by the BUILD awardees to upload, collect, store, and manage consortium participant data. The CEC Tracker assigns each participant a unique nine-digit numeric identification number. This allows the CEC to maintain longitudinal data regarding exposure to BUILD activities as individuals progress through their careers. The CEC Tracker allows authorized site administrators to add identifying elements to the CEC Tracker to assist with longitudinal tracking (e.g., site-level identification numbers). The CEC conducts quality review and risk assessment of the data. Access to the CEC Tracker requires authentication with a virtual private network (VPN) appliance in addition to CEC Tracker web application account verification. Because of the confidential nature of the data, the participant lists are not available for consortium or third-party use. Sites have ongoing password protected access to their own de-identified tracker data. Identifiable participant information is only provided to authorized educational officials at individual awardee institutions and is subject to their local IRB governance.

BUILD Institutional Records Data. Institutional Records (IR) data is essential for accurate tracking of student persistence and graduation, as well as faculty accomplishments (see Attachment 19). These data include (1) de-identified data for introductory science and mathematics courses, and (2) identifiable data for students and faculty who have provided consent through surveys. Awardees must use a secure file transfer service over an encrypted connection to transfer IR data. Identifiable site-level IR data is only provided to authorized educational officials at individual awardee institution and is subject to their local IRB governance.

As outlined in Section A.10, all data collected by the CEC is stored in a manner such that restricted information (e.g., name, address, contact information) is stored in a different system from study data such as survey responses. The restricted information is stored in a system behind the CEC firewall and operates on a private IP range. Only a limited number of authorized CEC staff are able to access these IP addresses. The study data is maintained in a separate system requiring authorized users using encryption. Any paper files used for data collection (such as handwritten interview notes) are stored in locked cabinets with access limited and controlled as with electronic data.

The Data Sharing Agreement for Phase II (see Attachment 25), most recently approved in September of 2019, was developed in conjunction with the awardees and the Executive Steering Committee. The agreement describes the requirements for data collection, integrity, storage, security, confidentiality, use, sharing, ownership, rights, and responsibilities.

B.2.2. Analysis Procedures

Qualitative Data. Observation and interview data from site visits and case studies will be analyzed in two cycles.[4] First, data will be assigned preliminary codes through attribute, structural, descriptive, and *in vivo* coding. During the second coding cycle, we will develop the categorical, thematic, and conceptual organization of the data. Through *pattern coding*, we will synthesize findings into more meaningful units of analysis.[5] By grouping similarly coded passages together and assessing the groupings for thematic commonalities, the final coding scheme will be established. Finally, through *elaborative coding*, we will examine the data with an eye toward the consortium level logic model (our conceptual framework). One of the drawbacks of a conceptual framework is that it may limit the inductive approach when exploring a phenomenon. To safeguard against becoming deductive, researchers will journal their thoughts and decisions and discuss them to determine if their thinking has become too driven by the framework. In qualitative inquiry, the researcher's values are not "controlled for" in the study design. Qualitative researchers use reflective journals as a strategy for examining personal assumptions and subjectivities. This reflexive practice provides transparency in the research process. Reflective journals will be used by our researchers to ensure that our process is inductive, that is, not overly reliant upon our conceptual program model so that our work may reveal contextually sensitive pragmatic descriptions of programs at the time of our data collection. Our researchers will discuss their reflective notes so to sharpen our insights and deepen our understandings of our observation and interview data. We will also be sure that our multiple data sources converge in an attempt to understand the overall case.

Quantitative data. Descriptive statistics such as counts, ranges, means, and frequency distributions will be employed using SAS and Stata software. Statistical methods to test our study objectives are described in Section B.1.2 above. A methods guide for conducting the statistical analyses along with potential research questions is provided in Attachment 26. Software used will be either SAS or Stata depending on the appropriate procedure.

B.3 Methods to Maximize Response Rates and Deal with Nonresponse

Multiple strategies are used to maximize response rates. First, all surveys are implemented with multiple modalities. The primary modality is online, but all surveys are formatted to be completed as computer-assisted interviews (conducted by CEC interviewers) and as paper-based surveys. Thus, respondents can choose the modality with which they are most comfortable.

Second, recruitment/retention strategies include non-monetary approaches known to improve response rates. These include providing respondents information about the contribution they will be making to the understanding of the important issues on which the project focuses[6] and having influencers send the introductory survey invitations.

Third, we have determined that successful recruitment and retention efforts require a monetary incentive. We are requesting to provide an incentive of up to $25 to each of our participants for

[4] Saldaña J. The coding manual for qualitative researchers (2nd ed.). 2014. Thousand Oaks, CA: Sage.
[5] Miles MB & Huberman AM. Qualitative data analysis: An expanded sourcebook. 1994. San Francisco, CA: Sage.
[6] Singer E & Ye C. The use and effects of incentives in surveys. Ann Am Acad Polit Soc Sci, 2013; 645: 112-141.

each survey they are asked to complete. Incentives are critical to our ability to recruit and retain a representative sample over the longitudinal follow-up period in order to track primary outcomes for the required program evaluation.[7] The choice of a $25 incentive is based on evidence from prior experimental work showing incentives in this range can improve response rates significantly[8] as well as the experience of members of our consortium with the value of such incentives in maintaining better longitudinal response rates.[9]

Finally, we will continue to emphasize to respondents as they enroll in the various programs how important the continued tracking of information is to the long-term evaluation of the program.

B.4 Test of Procedures or Methods to be Undertaken

All surveys are pilot tested with 5-9 respondents to ensure readability, flow, and time for administration. Semi-structured interviews are pilot tested with 1-2 respondents to ensure flow and time for administration.

B.5 Individuals Consulted on Statistical Aspects and Individuals Collecting and/or Analyzing Data

All plans for data collection and statistical analysis are the product of CEC investigators and staff, which includes Ph.D.-level biostatistics faculty as well as researchers with extensive expertise in program evaluation, with input from Consortium Working Group members and the Executive Steering Committee.

[7] Estrada M, Woodcock A, Schultz PW. Tailored Panel Management: A Theory-Based Approach to Building and Maintaining Participant Commitment to a Longitudinal Study. Evaluation Review, 2014; 38: pp 3-28.
[8] To N. Review of Federal Survey Program Experiences in Incentives. Bureau of Labor Statistics, July 23, 2015. http://www.bls.gov/cex/research_papers/pdf/Review-of-Incentive-Experiences-Report.pdf
[9] Estrada M, Woodcock A, Schultz PW. Tailored Panel Management: A Theory-Based Approach to Building and Maintaining Participant Commitment to a Longitudinal Study. Evaluation Review, 2014; 38: pp 3-28.