IDENTIFYING AND EVALUATING TRAINEE OUTCOMES

SUPPLEMENTAL REPORT

National Human Genome Research Institute
Advancing human health through genomics research

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

AAAS	American Association for the Advancement of Science
ACGC	Accreditation Council for Genetic Counseling
BEST	Broadening Experiences in Scientific Training
BUILD	Building Infrastructure Leading to Diversity
CERE	Center for Research Evaluation (University of Mississippi)
CMDB	Cell, Molecular, Developmental Biology, and Biophysics (Johns Hopkins)
CRTP	Clinical Research Training Program
CSULB	California State University Long Beach
CV	Curriculum Vitae
DART	Diversity in Addiction Research Training (Medical University of South Carolina)
DPAC	Diversity Post Doctoral Alliance (Johns Hopkins)
FIRST	Faculty Institutional Recruitment for Sustainable Transformation
FTE	Full Time Equivalency
GPA	Grade Point Average
HBCU	Historically Black Colleges and Universities
I3	Independent Investigator Incubator
IDP	Individual Development Plan
IMSD	Initiative for Maximizing Student Development
IRB	Institutional Review Board
MyNRMN	National Mentoring and Networking Platform
NASEM	National Academies of Sciences, Engineering, and Medicine
NHGRI	National Human Genome Research Institute
NIH	National Institutes of Health
NIH RePORTER	Research Portfolio Online Reporting Tools
NINDS	National Institute of Neurological Disorders and Stroke
NSF INCLUDES	Inclusion across the Nation of Communities of Learners of Underrepresented Discoverers in Engineering and Science
PI	Project Investigator
PII	Personally Identifiable Information
REDCap	Research Electronic Data Capture
REU	Research Experiences for Undergraduates
STEM	Science, Technology, Engineering, Mathematics

STPF	Science and Technology Policy Fellowship
TiDHE	Training, Diversity, and Health Equity
UMBC	University of Maryland, Baltimore
URM	Underrepresented Minorities
USDA	United States Department of Agriculture

Letter to the Reader from NHGRI

NHGRI is dedicated to cultivating a diverse pool of future researchers, clinician-scientists, healthcare providers, and educators. For many years, NHGRI has supported training and workforce development programs with a core commitment to enhance the diversity of the biomedical workforce. NHGRI's education and career training opportunities are done in collaboration with NHGRI's Extramural Research Program, Intramural Training Office, Education and Community Involvement Branch as well as external partners, such as professional societies. In 2021, NHGRI published an action agenda to enhance the diversity of the genomics workforce. This action agenda reflects a framework to connect early exposure to genetics, undergraduate and graduate efforts, and career development activities to careers in genomics and a greater diversity in the overall genomics workforce.

NHGRI also established a new office dedicated to Training, Diversity, and Health Equity (called TiDHE), which now includes NHGRI's Extramural training portfolio. In collaboration with the Divisions and Offices in the Extramural Program, the training team develops and supports initiatives that expand opportunities for genomics education and careers and cultivates genomics training programs and workforce development initiatives for all individuals, including those from groups underrepresented in biomedical research.

As part of these efforts, we have worked to develop an NHGRI toolkit to help programs all over the country learn more about best practices for evaluating training programs related to genomic workforce diversity, research, and genomic workforce programs. Effective training programs are crucial for trainees in achieving success in STEM. Studies have demonstrated that trainees are more successful when participating in training programs compared to those who do not. Thus, it is important to know whether a training program is effective in preparing its trainees for future biomedical and genomics careers. Additionally, understanding trainee outcomes is important because each stage of a trainee's career provides important milestones that should be met to ultimately have a productive and successful career in STEM. The NHGRI toolkit attempts to identify important success metrics that should be considered for all trainees in career stages ranging from undergraduate, graduate, postdoctoral, and genetic and clinical trainees, and even those that are pursuing non-academic career paths. Additionally, this supplemental report provides more in-depth information on each section within the toolkit including references, methodology, and lots of other details.

The hope is that this toolkit will be instrumental in supporting you in monitoring the success of participants in training programs you lead, support, or are a part of. For some of you, who are already deep in the process of implementing evaluation efforts, we hope this toolkit will complement your ongoing efforts. For others, who are newer to the evaluation process, or who are looking to reinvigorate your existing process, we hope that the toolkit will be a source of ideas. This toolkit should be regarded as a first effort in gathering information about evidence-based efforts related to trainee outcomes.

Here at NHGRI, the success of trainees really matters to us – we need to continue to grow a vibrant, successful, and diverse genomics workforce to be successful in all facets of genetics and genomics.

We hope you enjoy the content of this toolkit!

Thank you!

Background

[The National Human Genome Research Institute (NHGRI)](#) of the National Institutes of Health (NIH) is dedicated to cultivating a diverse pool of future researchers, healthcare providers, and educators. The Institute supports training and workforce development programs that are committed to maximizing diversity of the biomedical workforce. In 2021, NHGRI established a new office, dedicated to [Training, Diversity, and Health Equity (TiDHE)](#), to develop and support initiatives that expand opportunities for genomics education and careers; cultivate genomics training programs and workforce development initiatives for individuals underrepresented in biomedical research; and promote genomics research to improve minority health, reduce health disparities, and foster health equity. As part of this mission, TiDHE has worked to identify best practices in defining, identifying, and evaluating trainee outcomes within biomedical and genomics research. To share this information, a toolkit has been developed to guide training programs facilitated by NIH, other U.S. government agencies, foundations, industry, and academic institutions in developing streamlined approaches to monitor the success of former and current training program participants.

Purpose of This Toolkit and Supplemental Report

The goal of the NHGRI toolkit is to inform the development of streamlined approaches to identifying and evaluating trainee outcomes and measures of success for programs within genomic workforce diversity, research, and genomic workforce programs. The toolkit provides an overview of current methods, challenges, and best practices to consider when identifying and evaluating trainee outcomes and success in biomedical research. This report provides supplemental information to the toolkit for those interested in diving deeper into the information presented throughout the toolkit. Furthermore, a training program inventory was developed with a list of the programs reviewed as part of this project containing relevant program information.

Importance of Identifying and Evaluating Trainee Outcomes

Identifying and evaluating trainee outcomes of any program is important so that the overarching success of the program can be studied. Knowing the outcomes of a trainee allows programs to assess whether they are achieving their own program goals and to know if they are preparing trainees for their own career paths. This allows programs to utilize a feedback loop to constantly identify areas of improvement and success within their programs. The association of training programs with the successful outcomes of trainees has been widely studied in literature. Studies have shown that the quality of the training program, the mentorship and support provided to trainees, and the opportunities for research and professional development can have a significant impact on the success of trainees in biomedical fields (Lorden et al., 2011; National Academies of Sciences, Engineering, and Medicine, 2018). Training programs may also have a significant impact on the career trajectory of trainees from underrepresented population groups. Training programs that foster a culture of diversity, equity, and inclusion are associated with better outcomes for trainees from underrepresented groups (Hill et al., 2022). Biomedical training programs have the potential to address these issues by providing targeted support and resources for trainees from underrepresented population groups. Several studies have shown that interventions aimed at increasing the representation of underrepresented trainees in biomedical training programs have had positive effects on career outcomes (Campos et al., 2021; Odedina et al., 2022; Segarra et al., 2020). Therefore, to improve diversity, equity, and inclusion in the biomedical field, it is important to focus on the development and implementation of effective strategies to support and advance the careers of trainees from underrepresented population groups. If programs provide the necessary factors to ensure trainee success, the next step is being able to know whether trainees go on to pursue graduate degrees and continue into career paths in academia or elsewhere. As

trainees complete a program, identifying and evaluating trainee outcomes presents a huge challenge due to the time and effort it takes to find information, the lack of standard metrics to collect, maintaining the consistent engagement from trainees to provide their information, among other things. This creates problems for programs because it makes it difficult to assess whether a program is achieving its mission and goal and whether it is preparing trainees for their future careers. As a result, finding effective and efficient methods to evaluate trainee short-term and long-term outcomes can provide significant benefit to both programs and trainees.

Why does the toolkit matter and who is the intended audience?

Funding agencies and institutions need to have a thorough understanding of trainee outcomes, challenges, and key metrics that should be considered when developing and improving their trainee programs. This is essential because it helps program evaluation and improvement and ensures that future program participants are trained with evidence-informed strategies. Additionally, information presented in this toolkit can help programs understand what has worked in the past and common barriers to avoid to ultimately save both time and money. This toolkit goes beyond the traditional methods (e.g., surveys) for trainee identification and evaluation. It provides discussion on how to combine common methods with additional approaches for engagement and empowerment to effectively collect trainee outcome data suited for different career categories ranging from undergraduate and graduate students, postdoctoral fellows, students in genetic and clinical programs, and trainees pursuing non-academic career paths.

The information within this toolkit can further help program managers and directors, principal investigators (PIs), and mentors within agencies and other academic institutions to inform program policy changes and implement improvements to help programs and mentors understand disparities in trainee experiences and outcomes. Such policy changes can improve trainee outcomes and advance success. The goal is for interested parties to have a better understanding of things to consider when developing their own programs and the identifying and evaluating of their trainee outcomes.

Brief Methodology

Data collection necessary for the development of this toolkit was carried out in three components shown in **Figure 1**. A rapid review of peer-reviewed literature was conducted to gain insight into various training programs in the U.S. tailored for different career stages from undergraduate and graduate students, postdoctoral fellows, genetic and clinical trainees, and trainees who chose to pursue non-traditional career paths. A detailed description is provided in **Appendix A: Methodology: Literature Review**. Interviews were conducted with NHGRI intramural and extramural staff to better understand the Institute and how it evaluates its own trainees, as well as to identify best practices and barriers that it faces. These interviews then informed the development of questions for an additional interview with an NIH leadership representative to gain additional considerations for identification and evaluation of trainee outcomes at NIH. Full methodology for interviews, as well as specific questions asked, is provided in **Appendix B: Methodology: NIH Staff Interviews**. Twenty programs were selected for review that formed the development of the training program inventory, which can be found in **Appendix C: Methodology: Training Program Inventory Database**. This inventory provides a small representation of training programs that span various career categories mentioned above focusing on various disciplines. The data collected for this includes any publicly available information, as well as information gathered

through outreach with program managers and any program representatives that could be collected. It is not intended to be comprehensive or a full representation of all training programs.
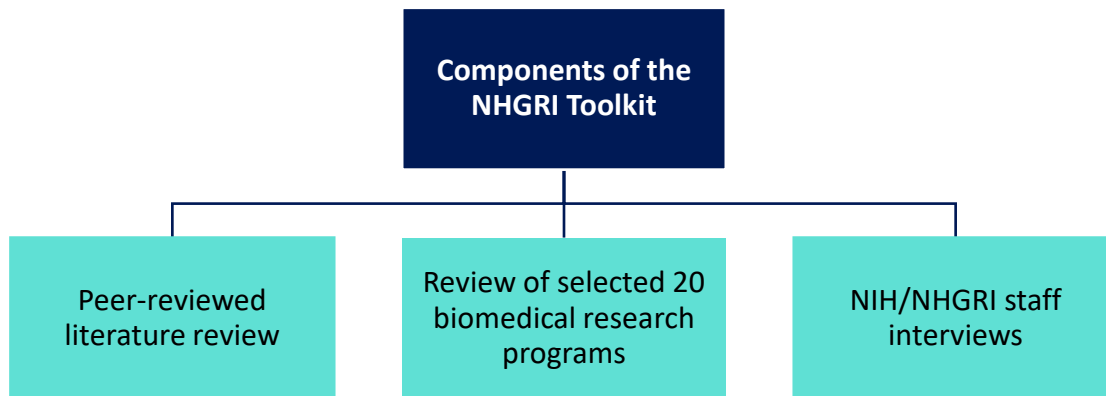
Figure 1. Research design

Data collection for all components of the NHGRI toolkit was guided by the following research questions:

- How do programs define successful outcomes for trainees?

- What are best practices for defining and tracking trainee outcomes?

- What are challenges and barriers to tracking trainee career development?

- How have programs addressed the challenges and barriers to tracking trainee career development?

Limitations

It is important to consider the limitations of this toolkit when trying to consider best practices, challenges, and any recommendations that are provided. The key limitations that should be considered include:

- **This is not a comprehensive review.** The components that made up the NHGRI toolkit **Figure 1,** were selective in approach and focused on gaining a representation of success and outcome metrics, best practices, challenges and barriers, and methods.

- **Six individuals representing NIH/NHGRI were interviewed for development of the toolkit.** The individuals interviewed for this project were primarily NHGRI staff and were not meant to be a representation of all training programs throughout NIH.

- **Different policies and procedures need to be considered when interpreting the findings and recommendations of the toolkit.** Agencies and institutions have policies and procedures that they need to abide by when implementing best practices or other changes suggested throughout the toolkit and report.

- **There are a large number of training programs that are a part of NIH, other federal agencies, and U.S. institutions that have their own approaches and metrics to evaluating trainee outcomes.** This toolkit is not intended to be a comprehensive review of all training programs but

rather a representative sample of select programs. The goal of this toolkit was intended to determine key best practices, challenges, methods, and innovative approaches to identifying and evaluating trainee outcomes.

- **Limited information is publicly available.** There were limitations in the amount of publicly available information collected to develop both the training program inventory and literature review. More specifically:

  - *The Training Program Inventory*: A list of information that we wanted to collect for each program was developed at the start of the toolkit development. Unfortunately, data collection for each program posed various challenges in the information that was publicly available. Some examples of this included program evaluation, methods to identify and evaluate trainees, as well as success metrics. Outreach was done to supplement the lack of information. Similarly, we could not reach representatives for all programs.

  - *Peer-reviewed literature:* Limited peer-reviewed studies focusing on trainees pursuing non-academic career paths was available and more longitudinal research is needed to understand the impact of training programs on non-academic career trajectories. More research is also needed to improve program components based on their association and applicability in non-academic career paths. Additionally, no studies were found that specifically curated their trainee outcome evaluation processes to meet unique needs of trainees from underserved populations.

- **Published information available for data collection methods are likely those that are commonly tracked.** Data collection methods and outcome measures were commonly available for training programs in academic settings. Unfortunately, these measures for alumni that opted for non-academic careers were not as readily available. Hence, the information presented in this toolkit does not provide sufficient information for trainees in non-academic settings or for evaluating outcomes not relevant to academic career trajectories. As a result, it is likely that methods found from review of literature are those that are more commonly tracked and may not account for novel methods that have yet to be analyzed and published.

Trainee Outcome Definitions and Success Measures

Most training programs define measures of success based on their mission and goals. There is no standard definition for trainee outcomes across career categories or within a category.

For this report, *trainee outcomes* are defined as the results, achievements, and/or accomplishments of trainees and alumni of a training program. These outcomes can include measures such as current positions, fields of work, job satisfaction, and the extent of application of knowledge and skills acquired during their training in their current role. *Measures of success* are the specific metrics used by programs to evaluate program effectiveness and program impact on trainees. These measures usually include publications, grants received, patents, promotions, time to receive tenure in academia, and other career-related milestones. Non-research related measures of success can also include job satisfaction, years of experience, application of genetics and genomics in the workplace, and mentorship and networking provided through the program. These metrics can help programs assess the value of the training program and identify specific program outcomes that are associated with shaping professional development of trainees and alumni.

Most training programs have a specific mission and goal. Program goals are informed by the career category and experience level of incoming trainees. For example, some programs such as the [University of Maryland Baltimore County (UMBC) Meyerhoff Scholarship program](#) and the [NIH Building Infrastructure Leading to Diversity (BUILD)](#) program prepare their trainees to pursue graduate school to obtain Ph.D.'s or master's degrees, while others throughout the NIH award grants, like the K01 award, which aim to prepare early career investigators to have independent research grants. Other programs like the [American Association for the Advancement of Science (AAAS) Science and Technology Policy Fellowship (STPF)](#) aim to prepare early career investigators with careers that connect science and policy. Program representatives we spoke with stated that trainee success outside of the program objectives is usually not considered. Thus, what success looks like can vary significantly for trainees and program administrators. For example, whereas obtaining a non-STEM degree might be considered an important career milestone by a trainee or alumni of a STEM training program, training programs may not consider that milestone as a success metric in their evaluation. This highlights the subjectivity in defining successful outcomes and emphasizes the importance for training programs to broaden the measures of success and develop ways to consider any career related milestones in their program evaluations.

Peer-reviewed studies on assessment of training programs and their impact on trainees, and research of training programs suggest that trainee outcomes and success can be broadly categorized into professional and psychosocial outcomes (**Table 1, Table 2, Table 3**). Since biomedical research programs generally seek to prepare trainees to have successful research careers, professional outcomes are more common metrics of success. Psychosocial and professional outcomes are discussed below.

Professional Outcomes

A review of our selected training programs found that most programs evaluate professional outcomes. Compared to professional outcomes described earlier, these are outcomes of trainees that may not consider some of the significant factors such as socio-economic status of trainees, research self-efficacy, mentorship, etc. Most undergraduate research training programs aim to prepare students for a future research heavy career. Studies have found positive association between participating in research training programs and participant success in a future STEM career (Wilson et al., 2018). For example, in the study conducted on Research Experiences for Undergraduates (REU) initiative at the National Science Foundation (NSF), data was extracted from 88 demographically similar students. The participants included 44 students who were actively a part of the REU program (participants) and 44 students who applied to the program but were not admitted (applicants). Five variables were assessed to determine success: field of study (STEM vs. non-STEM), highest degree pursued, number of scientific conference presentations, number of publications (peer-reviewed publications, research-based honors theses), and number of academic awards. To evaluate these measures, data was collected through PI tracking data, LinkedIn, scientific databases such as Google Scholar and PubMed, and other internet searches. The study found that more REU participants pursued Ph.D. programs when compared to applicants. The study also found that REU participants had more successful trainee outcomes than applicants as measured by higher instances of *publications, awards, and presentations*. These measures of professional outcomes are also common for graduate students as well as postdoctoral researchers. One of the most widely accepted measures of success in postdoctoral training programs is the number papers published in high-impact journals such as Nature, Science, and Cell, as well as the number of citations a researcher's work receives (Meyers et al., 2016; Ravenscroft et al., 2017; Springer Nature, 2019). Some programs may also put additional emphasis on authorship order, where first in the order is considered more valuable. For example, in an evaluation study of [NIH Broadening Experiences in Scientific Training (BEST)](#) awards, the authors reported increased research productivity that was measured by first-author research publications (Meyers et al., 2016). On the other hand, some authors

have reported that metrics such as the impact factor and h-index only measure the academic impact and may not report comprehensive impact of research (Ravenscroft et al., 2017). Furthermore, research skills, such as the ability to design and conduct a research project, problem solving and troubleshooting, institutional board review (IRB) submissions, analytical skills, and computer technology skills; and scholarship skills such as manuscript writing, authorship and integrity, dissemination of work, grant writing skills, and responding to reviewers' critiques are also important measures of success often included in program evaluation (Spence et al., 2018). Another measure of success in research is the amount of funding and number of grants a researcher receives. Some notable programs that include these measures in their evaluation metrics include the Indiana Clinical and Translational Sciences Institute's Independent Investigator Incubator (I3), NHLBI Programs to Increase Diversity Among Individuals Engaged in health-Related Research (PRIDE), Medical University of South Carolina's Diversity in Addiction Research Training (DART), and the Vanderbilt-Medical Scholars Program (Brown et al., 2018; J. Jones et al., 2022; Petrie et al., 2017; Rice et al., 2014; Spence et al., 2018). Grants include funding from government agencies such as NIH and the NSF, as well as non-profit foundations. These measures of success are also applicable to the training programs designed for early career investigators. For early career investigators, success in research can also be measured by a researcher's ability to mentor and train the next generation of researchers. Mentorship includes the number of high school students, undergraduate students, graduate students, and postdoctoral researchers that are mentored and trained, as well as the career trajectory of these individuals (Spence et al., 2018). Lastly, success can also be measured by the impact of research on the field and society. Impact of research is usually measured by the number of publications, grants, and patents but also the number of new therapies, drugs, diagnostic tools, and technologies that are generated by the research (Jorde et al., 2021). These outcomes divided by career stages are provided in **Table 1**, though these may vary with size, scope, and objectives of the programs.

Table 1. *Examples of professional outcome measures*

Career Stage	Examples of Professional Outcome Measures
Undergraduate Programs	• Field of study • Highest degree pursued • Graduation status • GPA • Academic and non-academic employment • Number of conference presentations, peer-reviewed publications, honors and awards, milestones, and accomplishments
Graduate Programs	• Graduation rate • Position post-graduation • Field of employment • Peer-reviewed publications • Grant funding • Patents and intellectual property

	• Use of skills learned in graduate school in current career
Postdoctoral Programs	• Field of employment • Position post-program • Mentorship • Peer-reviewed publications and presentations • Grant funding • Patents and intellectual property
Genetic counseling programs	• Completion of board exams and obtaining an academic/clinical job

Psychosocial Outcomes

In our assessment of publicly available information on trainee outcomes, we find that psychosocial outcomes are of interest to some undergraduate and graduate training programs as well as some clinical/medical and postdoctoral programs. Psychosocial outcomes consider a trainee's psychological, emotional, social, and cognitive well-being. Measuring psychosocial outcomes helps researchers capture a holistic understanding of program impact beyond assessment of educational and career outcomes. Additionally, assessment of psychosocial outcomes can help identify more effective program aspects critical to promote well-being and improving programs (Heck, 2004; Rossi et al., 2018; Weiss & Weiss, 1998). Psychosocial outcomes were more common in programs with undergraduate and graduate students as the intended participants. However, these outcomes were not as common for postdoctoral training programs. Review of postdoctoral programs found a lack of publicly available information around methods and successful practices to identify and evaluate postdoctoral trainee outcomes as well as program evaluation. This is important to note because researchers have often expressed the need for an efficient method to evaluate postdoctoral researchers' training outcomes to improve training (Polka et al., 2015).The time that trainees spend during their postdoctoral training is crucial in that it is a time for individuals to develop research competencies, and ultimately land competitive jobs (Levy, 2014). This stresses the importance of developing best practices and methods within postdoctoral programs to ensure that their trainees have successful outcomes.

The following table lists some examples of psychosocial outcomes that are considered in some of the training programs reviewed for the purposes of this toolkit.

Table 2. *Examples of psychosocial outcome measures*

Career Stage	Examples of Psychosocial Outcome Measures
Undergraduate and Graduate Programs	• Career satisfaction • Confidence in research • Community involvement • Research self-efficacy

Postdoctoral and Genetic Counseling Programs	• Career and professional development skills • Team building • Time management • Oral and written communication • Self-reflection • Cognitive assessment of leadership • Conflict and negotiation skills

Descriptions of psychosocial outcomes:

The descriptions for each psychosocial outcome below are based on peer-reviewed literature, review of selected training programs, and conversations with program managers and staff. However, different programs may choose to measure these outcomes differently depending on the scope of the program.

- Career satisfaction: The extent to which trainees and alumni exhibit satisfaction with their career based on the impact of the training program.

- Confidence in research: Trainees' belief in their own ability to successfully perform research tasks and activities in a laboratory and other research settings. These can include tasks related to designing and conducting research experiments, data analysis, data interpretation, and reporting results.

- Community involvement: Active engagement and participation of trainees and alumni with broader scientific as well as local communities to promote understanding of biomedical research and promote scientific communication.

- Research self-efficacy: Trainees' belief in their own ability to successfully perform research tasks to achieve desired outcomes. Self-efficacy plays a critical role in trainees' ability to connect and perform in training programs.

- Career and professional development skills: Understanding career options, self-reflection, networking, job search skills, team building, time management, oral and written communication, self-reflection, and cognitive assessment of leadership, conflict, and negotiation skills.

- Lifelong learning: Involves acquiring new skills, refining/updating existing skills, and expanding knowledge and understanding of various concepts of the field of research and career. It can include hands-on experiences with partners (e.g., internships and externships) outside of the university (e.g., biotechnology, science communication), or within the university (e.g., intellectual property, grants administration). The inference, in many cases, is that these partners may be potential employers of at least some of the trainees (Scalo & Freauff, 2020).

- Mentorship: Mentors can range from the primary research advisor of the mentee as well as peers. Mentorship networking process can also include connecting to alumni and professionals in their career of interest (Lenzi et al., 2020).

Non-Academic Career Paths

Recent data implies that graduate and postdoctoral candidates may not necessarily seek academic careers (Sauermann & Roach, 2016). In 2008, only 23% of biomedical research Ph.D. holders were tenured or on the tenure track and postdoctoral appointments declined by 3.6% from 2020-23 and by

2.2% from 2017-23 (NSF, 2020). Studies also suggest a decreasing interest of Ph.D. students in pursuing academic career paths, irrespective of race or ethnicity (Gibbs et al., 2014; Sauermann & Roach, 2016). These studies demonstrate the need to identify and evaluate trainee outcomes for those pursuing non-academic career paths. Interestingly, for most training programs we reviewed, there were no consistent criteria to evaluate outcomes for those trainees who opted for a non-traditional or non-academic career path. Additionally, many of the psychosocial and professional outcomes commonly considered in evaluation criteria are specific to academic career paths. This can make it difficult to identify clear career paths and define success outcomes for those choosing to pursue careers outside of academia. In fact, one study looked at training programs that provide business and experiential experiences to biomedical trainees and noted that providing training that prepares trainees for nonacademic careers "implicitly devalues the research-intensive academic career path" (Petrie et al., 2017). Such notions that value one career path over others or that limit exposure to training and information can make trainees feel as if their career choices are limited. Such an environment that values primarily an academic career path can be challenging for trainees who may want to pursue careers alternative to academia or may want to prioritize factors that are not common in academic careers, such as work-life balance and higher incomes. On the other hand, some training programs can influence faculty participants to become more supportive of non-academic career choices (Mathur et al., 2023). Literature highlights some outcome measures to consider for those pursuing non-academic career paths (Gibbs et al., 2014; Langin, 2019; Mathur et al., 2023; Sauermann & Roach, 2016). These outcomes are shown in **Table 3**.

Table 3. *Examples of outcomes measure for non-academic careers*

Career Stage	Example Measures of Outcomes
Non-Academic	- Product development - Intellectual property - Business growth - Corporate achievement - Awards and recognition - Patents - Other forms of intellectual property

Methods for Data Collection to Identifying and Evaluating Trainee Outcomes

Surveys

Program exit surveys and annual alumni surveys can be powerful tools for gathering program feedback. This data can help to measure the career trajectories of trainees and alumni and assess program outcomes.

- **Program exit surveys** are administered to participants immediately upon completion of the training program. Some of the measures commonly included in program exit surveys are satisfaction with the program design, instructors, and resources provided during the opportunity. Exit surveys have the benefit of providing timely feedback, essential for identifying

areas of improvement (Dugan & Komives, 2007). Program exit surveys can also be useful to provide insights into the professional goals of participants which can further be used to tailor programs to best meet participant needs (Carini et al., 2006). Surveys can also be used to identify specific skills and competencies that are essential for trainees' success as the field progresses over time.

- **Annual alumni surveys** are conducted every year after participants have completed a program. Annual alumni surveys are helpful because they provide longitudinal data on the career trajectories of alumni, enabling programs and institutions to monitor their professional progress and success over time. Annual alumni surveys can also help to assess long-term program outcomes (Pike et al., 2011). Surveys can help collect data to allow programs to explore statistical associations between their programs constructs and trainee career outcomes. Furthermore, when programs conduct long-term data collection through surveys, the differences reflected within the surveys can be determined over time ensuring that the program continues to stay relevant to the changing demands of the job market.

Interviews and Focus Groups

Programs such as the UMBC Meyerhoff Scholarship program and the AAAS STPF program reported meeting with trainees during the program and conducting interviews following program completion. Interviews and focus groups can provide rich information from program participants. For example, the dynamic interaction among participants in a focus group setting can help spur thoughts among participants and generate major ideas and insights to inform program improvement. Additionally, when interviews and focus groups are conducted with participants who are currently in a training program, participant feedback may inform small-scale program modifications. Furthermore, when conducted with participants who have completed the program, alumni insight may reflect a more holistic vision of the program, including what program components were more impactful than others, and what program components can be improved to better serve future trainees. However, there are limitations with this type of data collection. Depending upon sample size, it may be challenging to generalize information. Moreover, recruiting alumni, especially after a lot of time has passed since they finished the program, may prove difficult. Programs report limited engagement with trainees and from alumni after they have graduated from the program (Brown et al., 2018).

Professional Networking Platforms

The utility of professional networking platforms such as LinkedIn and ResearchGate expands beyond acting as a means for individuals to showcase their skills. These platforms offer valuable opportunities for program administrators to fetch data on the career trajectories of their trainees and alumni. Users of these platforms create profiles that include information about education, work experience, skills, and professional accomplishments such as grants and funding, and honors and awards. Additionally, program personnel can use these platforms to connect with alumni and conduct surveys and interviews to gather more detailed information about their work experience and career trajectory long after alumni have left the program. Programs need to be aware that a limitation of this approach is that only trainees who take the time to develop a profile can be found in this manner. Additionally, trainees may not keep their networking profile up-to-date or may not fully complete it. An interviewee mentioned this as a limitation that they have faced.

More information is provided on how programs can utilize user-managed databases in the section **Successful Practices.**

Emails

Another approach that PIs and mentors can use to check-in with alumni regarding their well-being and career progress and offer necessary mentorship or networking opportunities is through email. Occasionally connecting with alumni through emails can be useful in the long run to ensure better response rates in program related data collection. Email is also consistently used to provide trainees with program information such as updates, newsletters, surveys, opportunities, etc. Interviewees mention that sending email to trainees proves to be labor intensive. Other challenges that were mentioned were ensuring both consistent communication via email as well as keeping the most up-to-date email on file.

Evaluation

Evaluation, when conducted internally by program staff and externally by a third-party contractor, are common for programs given that funding agencies typically make program evaluation mandatory. Most evaluations employ more than one data collection method including but not limited to surveys, interviews, observations, focus groups, archival research, bibliometrics, and ethnography (Lenzi et al., 2020; Link & Vonortas, 2013; Patton, 2008; Rossi et al., 2018; Scalo & Freauff, 2020; Weiss & Weiss, 1998).

- **Formative evaluation:** Formative evaluation is conducted during program development and implementation to provide feedback and identify areas for improvement (Link & Vonortas, 2013). Formative evaluation is common for those training programs with uncertainty about the best program approach.

- **Impact evaluation:** Impact evaluation measures short-term, medium-term, and long-term program effects. This type of evaluation can help to measure the impact of the program on trainees' outcomes and career trajectories. It also investigates the associations between specific program components and observed results to ensure that the impacts witnessed are the results of the actual program components and not confounding factors (Gertler et al., 2016).

- **Process evaluation:** Process evaluation involves investigation of program implementation to explore best practices informing future implementation efforts (Durlak & DuPre, 2008).

- **Summative evaluation:** The type of evaluation conducted post-program completion to evaluate the overall efficiency, effectiveness, and impact of the program and to identify best practices for

future considerations (Patton, 2008). A basic summative evaluation approach would be to compare initial program objectives and goals with final program outcomes.

Others

Other methods to identify and evaluate trainee success that are not as common in the literature or reviewed training programs include secondary data collection or using methods that do not need advanced technical expertise. Some of these methods are discussed below.

- **Publicly available user-managed database platforms:** Program personnel can choose to extract publicly available information from pre-existing platforms, such as MyNRMN, ResearchGate, Google Scholar, PubMed, and NIH RePORTER. A major limitation with using these platforms is that only limited information on peer-reviewed publications and grants received by individuals are available. These platforms may not have the most recent information and may not reflect a complete picture of the career trajectory of an individual. Additionally, the impact and contribution of the training program on career trajectory may not be easily determined using these platforms.

- Using basic tools to develop program-level database: Some programs such as the NIH BUILD at California State University Long Beach (CSULB) have developed a database called the Annual BUILD Snapshot (Zuniga & Colbern, 2021). Using Microsoft Excel, this snapshot database collects personalized files that contain unique data related to academic and non-academic career trajectories, and some of the measures are graduation status, employment status, honors and awards, publications, conference presentations, milestones, and achievements for each trainee.

Table 4 discusses selected training programs and lists their use of different data collection methods to track, identify, and evaluate trainee outcomes for specific career categories. These programs were chosen based on data available from peer-reviewed literature.

Table 4. *Examples of selected training programs and their use of different data collection methods to track, identify, and evaluate trainee outcomes for specific career categories.*

Agency/ Program	Career Categories	Evaluation/Data Collection Methods	Outcome Measures
NSF-REU (Wilson et al., 2018)	Undergraduate students	- REU site principal investigator (PI) tracking data - Social media (i.e., LinkedIn and Facebook) - Scientific databases (i.e., Google Scholar and PubMed) - Internet searches	- Field of study (STEM vs. non-STEM) - Highest degree pursued - Number of scientific conference presentations - Number of publications - Number of academic awards
NIH-BEST (Meyers et al., 2016)	Graduate and postgraduate trainees	- Institutional-level evaluation - Consortium-level evaluation - Continuous quality analysis	- Trainees' ability to make informed career decisions - Short postdoctoral trainings - Institutional infrastructure

Packard Foundation-Hopps Scholars (Thompson et al., 2016)	High school graduates	• National Student Clearinghouse • Data collection by the Provost's office • Program-level data from Hopps Program PI	• Graduate program enrollment (in any field) • Graduate program enrollment in STEM • Health–related degree program enrolment • Carnegie classification of institutions attended by the Hopps scholars
NIH-NRMN (Ahmed et al., 2021)	Any	• Network connections • Profile transition changes	• Intra- and Inter-institution network connections • Peer mentoring connections
NHLBI-SIPID (Rice et al., 2014)	Early career investigators/ junior faculty	• Program evaluation • Program effectiveness • Mentees' progress over time	• Grants and publications under development, submitted, funded/ accepted/published
NHLBI-SIPID and NHLBI-PRIDE (Rice et al., 2017)	Early career investigators/ junior faculty	• Program evaluation • Online surveys	• Promotions • Tenure • Publications • Grant submissions and funding secured • Honors and awards • Presentations • Professional memberships • Election to offices • Serving as a journal/grant reviewer
NIDA-DART (Jones et al., 2022)	Postgraduate Year 4 and Year 5 residents	• CV of trainees and alumni • Online search using PubMed, NIH RePORTER, and Grantome • Program staff reaches out to trainees and alumni annually	• Abstracts presented or accepted for presentation at scientific meetings • Manuscripts published or in press • Attainment of research, career development, teaching, or other awards
NIH-CRTP (Ognibene	Medical and dental students	• Database of CRTP participants	• Percentage research time

et al., 2016)		• Follow-up surveys • Latest version of trainees' CV • NIH RePORTER and PubMed	• Extramural funding • Peer reviewed publications • Additional degrees obtained
USDA (Dockry et al., 2022)	Individuals who have completed a qualifying advanced degree (e.g., masters or professional degree) within the past two years	• USDA Human Resources Management stores data for all interns/ trainees and provides publicly available data [USDA Forest Service employee demographic data for diversity and inclusion analysis, 1995-2017]	• Individual Development Plan (IDP) • Transition to permanent employment • Career trajectory metrics such as length of service, career advancement, initial pay grade level, and separation type
NIH BUILD (CSULB) (Zuniga & Colbern, 2021)	Undergraduate students	• Annual BUILD snapshot	• Academic and non-academic career trajectory • Graduation status • Employment status • Honors and awards • Publications • Conference presentations • Milestones and achievements

Successful Practices

Dedicate Personnel to Oversee Identification and Evaluation of Trainee Outcomes

Programs employ designated personnel to collect data related to trainee and alumni outcomes, to develop and/or maintain databases, and to maintain communication with trainees and alumni.

Typically, programs hire or designate personnel to evaluate the impact of the program on trainee outcomes. The NASEM Air Force Science & Technology Fellowship Program has coordinating staff to oversee data collection of trainees through emails and surveys. The UMBC Meyerhoff Scholars Program has a dedicated evaluation team whose tasks include supporting trainees throughout the program, developing community building, and maintaining connection with trainees following completion of the program to help ensure success. Dedicated staff members in training programs tend to be responsible for many tasks including annual collection of CVs, building publicly available snapshots, developing, and collecting information for institutional database folders, as well as establishing a system for overseeing the development and management of trainee databases (Zuniga & Colbern, 2021). More discussion on development and maintenance of institutional/program databases is provided in section: **Develop and Maintain an Institutional/Program Database**.

Review of training programs uncovered that program personnel tend to maintain consistent communication and engagement with alumni as a best practice. Earlier methods of alumni engagement included regular emails to ask for updates or provide new relevant program information. However, over time, several programs have gone beyond this type of basic rapport building and may engage in more involved approaches to engaging with alumni, such as leveraging web portals or social media. Consistent engagement from familiar faces is useful in maintaining communication. For example, AAAS Mass Media Fellowship Alumni LinkedIn group is a relatively active LinkedIn group with more than 120 members. The BUILD CSULB program also has an active LinkedIn group of more than 350 members. Similarly, AAAS has an active Twitter account with more than 15,000 Twitter followers. Though there are several automated options to run social media, a team dedicated to relationship building, social media coordination, program outreach, and data collection is needed to constantly collect new trainee and alumni information and consistently update databases. A limitation that program staff need to consider are those trainee profiles that may be out of date or incomplete. Staff may need to find alternative approaches to data collection in these situations.

Develop and Maintain an Institutional/Program Database

When properly designed and maintained, program-specific databases can provide valuable insights into program success as it relates to trainee career success.

Institutionally managed databases include those maintained by institutional human resources or program administrators. These databases offer a systematic way to collect and manage data curated to program needs and objectives regarding evaluating trainee outcomes. Program administrators can analyze trainee career trajectory data to identify trends and patterns in training and alumni career outcomes, assess program effectiveness, and enhance informed decision making. Additionally, institutions can have a central location where program personnel regularly update information to ensure accuracy. These databases also offer several benefits that help address limitations discussed in section: **Barriers to Identifying and Evaluating Trainee Success.** Institutionally managed databases offer a secure location with enhanced security as the database is managed and maintained by the institution itself. These databases also allow program administrators to comply with their organization's privacy and security policies, ensuring that sensitive information is safeguarded from unauthorized access or misuse. These benefits can help address concerns regarding trainee privacy which will be discussed later. Some examples of institutionally managed databases include AAAS Fellows Central which gathers data from trainees that are or were a part of the AAAS STPF. Similarly, NIH also maintains the NIH Trainee Alumni Database. Databases can be designed to allow for easy updates and modifications, and tailored monitoring of career progress of trainees and alumni. A limiting factor of institutional databases is the cost of development and maintenance. Databases that are tailored to program needs require experts in developing databases. Additionally, the cost of program personnel responsible for ensuring optimization and accurate data collection needs to be considered.

An alternative to the development of tailored databases such as AAAS Fellows Central and NIH Trainee Alumni Database is to use database programs that have been developed prior and can be populated with fields of interest. These databases offer training programs with the ability to use a template to build a database that fits their needs. An example of this type of data is discussed below:

REDCap is a secure web-based application designed to support data capture for research studies. It allows programs to create customized surveys and databases for evaluating trainee career progression in relation to the scope of the training program, including educational experiences and professional development activities (Harris et al., 2009). REDCap has extensive data collection and management capabilities to conduct longitudinal analysis and evaluate program effectiveness. REDCap can be difficult

to use for a beginner and may present data storage limitations for large long-term databases (Harris et al., 2009). Automated integration of data collection and utilizing open-sourced databases can aid in evaluation of trainee outcomes when staff and funding resources are limited.

Open-sourced databases include those that have been previously developed and are made freely available for use. In the absence of adequate resources, these platforms, while not a standard method of data collection, can be used to obtain limited types of information such as peer-reviewed publications and government grant funding. Some examples of platforms from which information can be obtained are discussed below:

MyNRMN is an online platform that facilitates informal networking for researchers and trainees. This platform can help institutions gather data on the trainees' and alumni's career development through mentorship activities and professional connections (Ahmed et al., 2021; Javier et al., 2021; Sorkness et al., 2017). MyNRMN presents several limitations to collecting data. This platform is effective if users keep their information regularly updated. Only high-level data such as current position and past work experience of the users are generally available. The database may not have information on common professional measures such as grants, peer-reviewed publications, awards, and accolades etc.

NIH RePORTER and **Grantome** online database and reporting systems collect data related to research grants contracts, and publications. This platform can be used to identify trainees and alumni who have secured research funding and analyze the type and scope of research projects. One limitation of NIH RePORTER and Grantome is that the platforms are limited to provide insight into research trajectories of trainees and alumni in research-intensive fields, given the nature of the data that these systems collect.

ResearchGate, PubMed, and **Google Scholar** can be used by program personnel to obtain information on peer-reviewed publications of alumni. However, these platforms are not comprehensive and do not provide any other significant information regarding career trajectories of trainees and alumni.

Commit to Data Transparency with Your Trainees
Programs need to ensure that they are completely transparent with how trainee data will be used.

There are many challenges when trying to obtain information from trainees and alumni. Trainees and alumni may hesitate to share data because they do not know how the data will be used, stored, and whether it will be anonymous to the public when published. As a result, trainees may not be willing to share information, a problem which may be more pronounced following their completion of the program.

NIH has guidelines around data transparency that may help in such data collection efforts. At NIH, transparency is discussed as follows: "NIH should be transparent about information policies and practices with respect to Personally Identifiable Information (PII), and should provide clear and accessible notice regarding creation, collection, use, processing, storage, maintenance, dissemination, and disclosure of PII" (NIH, 2017). Furthermore, NIH states that "privacy is an essential freedom. It is the right of individuals to determine for themselves when, how, and to what extent personal information in the possession of NIH and its contractors is communicated to others" (NIH, 2017). A study done with the CSULB BUILD training program identified that being transparent with respect to data collection and its purpose was one key factor that can aid in alumni data collection (Zuniga & Colbern, 2021). Considering this information, training programs may wish to provide trainees with a rationale for sharing information and how the data will be used for program improvement.

"We have 20% no response rate for some stuff I believe because of distrust or other barriers. So, I think the more we talk about the importance of collecting trainee outcome info and how it is utilized to trainees, this will help get higher response rates. Especially with regards to demographic data as many are unsure what they're being used for." -NIH program director

Additionally, programs should be completely transparent and discuss how the data shared will inform program and policy changes and strategic plans. Furthermore, being clear that data can help programs understand disparities like pay-gaps (Schaller & Stalsburg, 2022), representation in different workforces, and at different career stages ultimately helps trainees understand the various changes that are implemented within the program. The goal is to help trainees and alumni understand how they and the programs can benefit from sharing outcome data.

Perform Regular Program Evaluations

Program evaluations are essential to monitor and assess the changing professional trajectory of program training and alumni.

Program evaluations allow programs to monitor their own programs to identify best practices and areas for improvement. These evaluations may include assessment of trainee satisfaction, acquisition of skills and competencies, and post-program career success (Khandker et al., 2009). Systematic evaluation of outcomes enables institutions and programs to identify patterns and trends in professional career trajectories and make data-informed decisions about program improvements and resource allocation (Newcomer et al., 2015; Poister, 2015). Evaluations can use one or a combination of data collection methods and are usually required by funding agencies (Fetterman & Wandersman, 2005; Khandker et al., 2009; Newcomer et al., 2015).

Internal evaluation: Internal program evaluations are conducted by the institution or the program themselves. Conducting internal evaluations, which may include activities such as exit interviews, are effective approaches for monitoring trainee achievements during and immediately after the program. There are several benefits of internal evaluation. For example, internal staff, in comparison with external evaluation teams, often have a deeper understanding of the program's concept, design, and context; increased ownership of the evaluation process; and more of an ability to make real-time improvements (Fetterman & Wandersman, 2005). Internal evaluations provide opportunities to tailor the program while the evaluation is still in process. They also provide opportunity to address training concerns (Newcomer et al., 2015). Finally, internal evaluation is better suited in programs that are new or when a program is being significantly modified. They might be preferred if the program budget does not allow for an external evaluation.

External evaluations: External evaluations offer a non-biased approach to assessing program outcomes. An external perspective can be helpful to identify issues that may be overlooked by internal evaluators due to their familiarity with the program (Hatry et al., 2015). External evaluations often are associated with a higher degree of credibility and legitimacy especially when it comes to funding and stakeholder support (Hatry et al., 2015; Rossi et al., 2018). Johns Hopkins Cell, Molecular, Developmental Biology, and Biophysics is an example of a program that regularly conducts external program evaluations. For the Johns Hopkins program, trainee outcomes of interest include graduation rates, employment fields, publications and their impact factors, current position, salary, field/industry, career satisfaction, use of skills used in the program, and satisfaction with the program. Most sites that are a part of the NIH BUILD

program also hire third party evaluation teams. For example, NIH BUILD CSUN works closely with the Center for Research Evaluation (CERE) for their evaluation needs. Some of the measures assessed for the BUILD CSUN program include undergraduate training retention, graduate/post-baccalaureate education after the program, research self-efficacy, understanding of research ethics and the importance of diversity, confidence about research abilities, campus community belonging, and likelihood of pursuing a scientific-related career.

Engage with and Empower Trainees

Trainee engagement and empowerment techniques can make common successful practices even more effective.

As described in the section **Barriers to Identifying and Evaluating Trainee Success**, maintaining trainee engagement to ensure the consistent sharing of their outcome data is a challenge to evaluating trainee outcomes. Additionally, interviewees mention that ensuring the trainees comply with requirements and getting them to provide information can be a challenge, an observation which is supported by a study of the CSULB BUILD program (Zuniga & Colbern, 2021). An increased sense of engagement and empowerment among trainees and alumni who are being asked to provide information for evaluation purposes may help address this barrier. Outreach with training programs suggested a few best practices to help trainees feel a sense of engagement and empowerment. These include techniques shown in **Figure 2.**

Figure 2. Trainee engagement and empowerment techniques to address limitations regarding low response rate from trainees and alumni

Community Building

Community building with trainees throughout the duration of the program has been noted as a successful practice. Interviewees mentioned that building community among trainees and creating a sense of belonging to the program helps trainees want to contribute to its success. Approaches to community building, as reflected by the review of training programs for this study and by published literature, include the following.

- Providing opportunities for social interaction among trainees and staff throughout the program. The NIH BUILD program mentioned that they have dedicated staff members personally reach out to trainees to have conversations regarding the program and collection of required information such as surveys. The AAAS STPF has dedicated staff assigned to specific agencies and trainees. Staff members set up quarterly check-ins to discuss progress and to see how trainees feel about their placement within the fellowship.

- Providing the necessary resources to support trainee success. Such resources include information about relevant opportunities, support for developing networks, and other resources to help prepare trainees for success throughout the program. The Meyerhoff Scholarship program provides such resources to its trainees, for example by helping to facilitate collaboration and support among trainees. Additionally, the program has its Summer Bridge Program, which aids trainees in transitioning from undergraduate to graduate school. These opportunities are highlighted in this video, which provides testimonials from Meyerhoff fellows.

- Recognizing trainee achievements through various outlets. The AAAS STPF utilizes LinkedIn, blogs, and their podcast to highlight many of their current and previous fellows' accomplishments. This includes sharing fellows' posts, highlighting current and previous fellows, and conducting interviews with current and past fellows.

- Creating networks of support by facilitating informal virtual or in-person networking events. Many programs support and advertise both formal and informal gatherings with current and former trainees as well as with mentors through virtual or in-person meetups. This allows trainees an opportunity to discuss topics and ask questions that might not normally come up through day-to-day tasks.

- Establishing strong alumni/trainee networks by inviting alumni to speak with trainees, enabling prolonged program connections with alumni. The NINDS Summer Internship Program has put together webinars in which previous trainees discuss their career path and share currently available opportunities. AAAS STPF puts on a yearly career summit where former fellows congregate to discuss resources or potential job opportunities available within government, non-profit, and other organizations with current fellows. Additionally, many programs such as AAAS STPF, NIH BUILD, and the Meyerhoff Scholarship program have developed LinkedIn groups that connect trainees with alumni, mentors, and program staff.

Maintaining Mentor-Mentee Relationships

High-quality training programs that provide hands-on experience, opportunities for research, and mentorship from experienced scientists have been found to be associated with better outcomes for trainees (Hund et al., 2018). Similarly, mentorship and support from experienced scientists have been found to be associated with the success of biomedical trainees (Barker, 2006; Eby et al., 2008; Hund et al., 2018). Studies have shown that trainees who receive mentorship and support have better outcomes, including higher rates of publications, grants, and advancement in their careers. The importance of mentorship in evaluating trainee outcomes was further emphasized through interviews with NIH and NHGRI staff. Establishing staff-trainee, PI-trainee, or trainee-alumni mentorship pairings can be

beneficial to programs because it helps trainees feel a sense of belonging and ultimately leads to more successful outcomes (Mentor Collective, 2020; Wright-Mair, 2020).

Programs and agencies have suggested several ways to help establish mentorship. These include:

- Incentivize university mentors to continue rapport building with trainees and alumni. Various grants that are awarded by NIH incorporate mentorship as a required component of the program, such as NIH's Faculty Institutional Recruitment for Sustainable Transformation (FIRST) program and the Initiative for Maximizing Student Development (IMSD) program. Similarly, the NSF Inclusion across the Nation of Communities of Learners of Underrepresented Discoverers in Engineering and Science (NSF INCLUDES) program has awarded grants to some programs that have mentorship as a key component. Another effort to strengthen mentorship includes reducing faculty teaching hours to accommodate mentorship efforts like the NIH Independent Investigator Incubator (I^3) program. This program is an early career training program in which I3 Mentors are compensated with 5% Full Time Equivalency (FTE) for each mentee, and can go up to 15% (Spence et al., 2018). Outreach with the NIH BUILD program suggested that sending personalized emails from familiar staff members to check in on trainees was a successful best practice.

- Establishing and encouraging opportunities for staff and alumni to serve as mentors for current and future trainees. The NIH BUILD program encourages its staff to establish connections with trainees and to maintain these connections throughout the duration of a trainee's participation in the program. The Meyerhoff Scholarship program also encourages its alumni to become mentors. They have developed a dedicated website for alumni to volunteer to become a mentor.

> *"[On adding other metrics] Ideally, I would like to have this really big community of alumni who are part of this network, who would be available as a mentor for the younger generation or the early career [researcher]. That's not really an outcome, but [a best practice]"* -NHGRI program director

Incentivizing Trainees/Alumni and Mentors to Share Data
Through outreach and review of selected training programs, it was found that incentivizing and, in certain situations, requiring the sharing of information was successful in getting trainees to share data. Many programs such as NASEM's Science and Technology fellowship, the AAAS Science and Technology Policy Fellowship (STPF), and the ORISE research program require their fellows to fill out exit surveys or participate in interviews following program completion. Additionally, the AAAS STPF requires participants to partake in surveys in order to successfully complete the program and receive alumni benefits, including access to the alumni portal and alumni resources through AAAS STPF Fellows Central.

Other approaches that are demonstrated to incentivize trainees/alumni and mentors to share data is educating trainees on the importance of surveys and how program success reflects on their own success. Additionally, showing trainees that program feedback can help improve the program and provide better outcomes for future trainees can encourage trainees to provide feedback.

Incorporating Feedback Loops
Continuous program improvement is key to ensuring program and trainee success. Utilizing data provided by trainees can give program administrators a starting point for improvement. Including survey

questions about trainee satisfaction with the program allows program administrators to learn about what is working well and potential gaps in their programs. Empowering trainees to provide feedback on all aspects of the program, including evaluation of trainees, can be effective in keeping them engaged. This best practice helps programs with their own internal improvement but also demonstrates to trainees that they provide value to the program and develops a sense of belonging.

Maximizing Effectiveness through Combination of Best Practices

Many programs reviewed utilize several practices mentioned above, such as hiring dedicated personnel, using databases to collect information, and performing or sponsoring program evaluations. Outreach with training programs, interviews, and literature uncovered the idea of engaging and empowering trainees as an additional method that can support efficient data collection for evaluation purposes. Combining these approaches to obtain the most information from trainees may be more effective than any one approach by itself. When trainees feel engaged and empowered, the more common techniques prove to be more successful. This approach is utilized by both the Meyerhoff Scholarship program and the NIH BUILD program, two programs that have demonstrated the ability to prepare their trainees to have successful outcomes following program completion (UMBC, 2022; Yomtov, 2013).

Barriers to Identifying and Evaluating Trainee Success

Lack of Time, Effort, Funding, and Resources

Data collection as well as maintaining and regularly updating trainee information may require significant time, effort, funding, and other resources.

Identifying and evaluating trainee success to explore the impact of the program on trainee careers can be labor intensive. As mentioned earlier, many resources may be needed to collect accurate information; maintain consistency in the collected information; establish databases to store, maintain, and protect the collected data; and analyze and use collected data to inform training program changes.

Implications

- **Limited capacity to design databases:** With limited resources, a comprehensive and adaptable evaluation system that accommodates the needs of biomedical training programs may be challenging to design (Sauermann & Roach, 2016). Staff training is also crucial in implementing and effectively using evaluation systems. Dedicated time and resources are needed to ensure the staff is competent in using the system to facilitate program improvement.

- **Limited datapoints:** Using a single source of data can be ineffective in identifying and evaluating trainee success. In some situations, developing a robust evaluation system requires integrating data from multiple sources. However, insufficient resources may be a limiting factor in data integration. Automated techniques such as machine learning or data analytics could enhance the efficiency of an evaluation system and data collection. Lack of consistent funding to support programs can also result in inconsistent data acquisition on trainee progress (Lauer et al., 2017). Usually, the cost of hiring personnel with expertise in specialized data collection techniques may be higher than the program capacity.

Lack of Standardized Metrics Collected

There are no standard definitions for trainee outcomes or measures of success across different training programs.

As described earlier, most training programs collect information around professional outcomes and in some instances psychosocial outcomes as well. One issue that was apparent throughout literature and through review of training programs was inconsistency in collected information that centered around measures of success. This issue can potentially be attributed to a lack of standard procedures for data collection and maintaining trainee databases. This poses an obstacle to comparing different programs or when determining which specific program outcomes are associated with training success.

Implications:

- **A lack of standard definitions for trainee outcomes and success makes it difficult to compare data across programs:** Metrics used to determine and assess trainee outcomes and success vary among programs. Absence of standardized definitions has resulted in a wide variety of reported outcomes across different biomedical research training programs.

- **Confounding variables:** Evaluation studies from programs such as the NIH BUILD that aim to improve representation of individuals from underrepresented population groups consider factors such as race, gender, ethnicity, and mentorship in their analyses (Saetermoe et al., 2017; Yomtov, 2013). However, other factors such as socioeconomic barriers, limited access to information, lack of mentorship and networking opportunities, bias, and discrimination were infrequent measures of evaluation that was found in review of other training programs.

- **Inefficient resource allocation:** In the absence of standardized metrics of identification and evaluation of training outcomes, it can become challenging for funding agencies to assess which programs require funding support or which programs may yield better returns on investment (Fuhrmann, 2016; Lauer et al., 2017). Additionally, lack of standardized definitions can hinder identification of best practices across different training programs or the benchmarking of program performance against a national standard, thereby hindering program improvement.

Inconsistent Trainee Engagement

Maintaining consistent engagement with trainees during and beyond training programs is a significant challenge reported by programs.

When investigators rely on reaching out to alumni without having previously maintained rapport with trainees and alumni, the response rate for data collection efforts such as through emails and surveys is not effective (Brown et al., 2018; J. Jones et al., 2022). Many alumni may not stay engaged with the organization after completing a program, which can make it difficult to identify and evaluate their accomplishments and progress (Cuadra et al., 2019). This can be particularly challenging for organizations that rely on self-reported data from alumni. For example, the NIH-BUILD program team at CSU-Long Beach tested a Program Database Management system where all information was collected by BUILD mentors who further relied on informal relationships such as LinkedIn (Zuniga & Colbern, 2021). When this approach became unscalable due to a growing number of alumni, the team started sending alumni an annual Qualtrics survey for more direct data collection from alumni. However, the response rate to the survey was low, and the program team identified that the lack of continued rapport contributed to this low response rate (Zuniga & Colbern, 2021). Similar to these findings from the literature, interviewees also mentioned that getting trainees to provide information is a challenge, noting that trainees seem to be too busy to share information. Interviewees also noted that programs try to address this issue by repeatedly reminding trainees why sharing information is important.

Implications:

- **Inability to provide support throughout the program can lead to reduced trainee satisfaction:** An absence of adequate infrastructure and technology can hinder program personnel in their ability to provide support and guidance to trainees. This lack of support results in a decrease in trainee satisfaction with the program and as a result, lowers the success rate when collecting trainee outcome information (Feldon et al., 2019).

- **Lack of consistent mentorship:** Lack of consistent mentor-trainee engagement could result in lost communication between the mentor and the trainee, and can pose a challenge to career and skills development for trainees that could have benefitted from mentorship (Byars-Winston et al., 2017; Feldon et al., 2019). Inconsistent mentor-trainee engagement also characterizes as an example of failed mentoring relationship, often associated with competing interests between the mentor and mentee (Straus et al., 2013).

- **The most updated and recent information may not be publicly available:** In the absence of necessary funding and resources, program personnel can rely on publicly available information for secondary data collection. However, evaluating trainee outcomes and success relies heavily on the assumption that publicly available information on platforms such as LinkedIn and ResearchGate provide the most recent and correct information about trainee career trajectory.

Trainee Privacy Concerns

Trainees may be hesitant to share data because of either lack of understanding of utilization of information or fear of misuse of information. Studies demonstrate that society has concerns about collection of personal data and how it is used (Hsu et al., 2022).

Implications:

- **Fear of misuse or having a lack of knowledge regarding data utilization:** Trainees may avoid sharing data due to concerns regarding potential misuse of personal information such as demographic data, which could exacerbate their risk of facing discrimination or unintended bias (NIH, 2012). This concern can result in inaccuracies and limit how much information trainees and others may make available to training programs, networking sites, and other databases that collect information. Overall, this poses a barrier for programs that would like to collect the most accurate and up-to-date information when identifying and evaluating trainee outcomes.

- **Fear related to data security and privacy breaches:** Because of issues related to data security, information breaches, and cybersecurity attacks, a trainee may hesitate to share their outcome data. They may fear the possibility of data breaches leading to unauthorized access to their individual information (L. Jones & Wells, 2007).

Restricted Data Sharing Policies and Regulations

Policies and regulations related to what trainee information a program can collect frequently change, especially demographic data and PII.

While collecting data for evaluation purposes, training programs often must adhere to relevant federal, state, and/or institutional policies relating to data collection, data storage, data security, and protection of trainee privacy. These restrictions may be particularly important for training programs that are federally funded which can limit what information can be collected. Furthermore, the agency in which the training program is a part of may have their own mandates that they must abide by when trying to collect PII. More specifically, interviewees mentioned that gaining clearance from the Office of

Management and Budget (OMB) can be a challenging and lengthy process. Overall, this can limit the information required to identify and evaluate trainee outcomes.

Implications:

- **Data collection and reporting:** Policies, regulations, and restrictions related to what trainee information can be collected by programs are put in place to protect privacy and confidentiality. At the same time, these regulations can limit the ability to collect essential information required to assess program impact and trainee success. These restrictions also limit the ability of various programs to compare outcomes. Additionally, it may become challenging to identify best practices to identify and evaluate trainee outcomes suited for specific population groups. The absence of certain data, such as those restricted by policies and regulations, can pose an obstacle to investigating long-term trainee progress and in identifying program components and/or patterns that specifically contribute to desired milestones throughout trainees' careers.

Components of an Ideal System for Identification and Evaluation of Trainee Outcomes

An ideal system suitable for identification and evaluation of trainee outcomes may incorporate several successful practices. These practices are discussed below.

Develop a regularly updated database to store and maintain trainee data: As discussed in the section: **Develop and Maintain an Institutional/Program Database**, databases provide a location that can allow programs to manage and store data on a consistent basis (e.g. annually, biannually). Analysis can also be done to inform program changes and for program and evaluation personnel to investigate associations between program components and trainee outcomes.

Ensure that current and past trainees feel connected, empowered, and engaged: Consistent networking and engagement with alumni beyond the trainee program may support higher response rates for data collection. Consistent engagement may also ensure continued support and mentorship for trainees and better response rates for programs. These topics are further discussed in section: **Engage with and empower trainees**.

Evaluate program success and implement necessary changes: Conducting program evaluations can be helpful in understanding how effective a program is in preparing trainees for their future careers. Additionally, findings from evaluations can be used to identify which program components need to be introduced, improved, revised, or removed. More information on program evaluation can be found in section: **Perform regular program evaluations.**

Additional Considerations

As described in the toolkit, there are many methods to identify and evaluate trainee outcomes, as well as documented successful practices to incorporate when designing and implementing trainee programs. A combination of more than one practice is suggested to successfully evaluate trainee outcomes.

In addition to implementing successful practices, there are some additional ideas that programs may want to consider when developing a system for assessing trainee outcomes:

- **Trainee Well-Being and Community Building.** Investing in trainee well-being and community building supports consistent trainee engagement that, in turn, facilitates successful identification and evaluation of trainee outcomes. It is important to recognize that trainee success and program success may often be different. It is essential for the program to emphasize that individual trainee success is valued and recognized.

- **Varied Measures of Success.** In the process of investigating the association between program components and trainee outcomes, it is essential to account for those factors and milestones that may not directly align with program goals. For example, even though one of the program goals may be to prepare trainees for tenure-track careers, programs should still consider whether a non-academic career path can be measured as an indicator of success.

- **Regular Program Evaluations.** Training programs can benefit from regular evaluations as much as other types of programs. Evidence-informed evaluation procedures that are conducted at various stages of program implementation (before/during/after) allow for comprehensive data collection. Additionally, measures such as psychosocial and professional outcomes are suggested for data collection to allow for a more comprehensive and reliable data analysis process.

- **Success Measures May Need to be Tailored, and New Measures May Need to be Developed.** Programs must consider factors that disproportionately impact individuals from underrepresented population groups. Research has shown that individuals from underrepresented groups in the biomedical workforce, including individuals from racial/ethnic minorities and individuals with disabilities, often face barriers to career advancement and are underrepresented in leadership positions (Alegría et al., 2019; Ilonze et al., 2022; Webb et al., 2022). Individuals from underrepresented population groups in the biomedical workforce are also less likely to receive promotions and may have slower career progression compared to those who are not from underrepresented minoritized groups (Allen-Ramdial & Campbell, 2014; Carroll et al., 2022; Zambrana et al., 2023). Several factors have been identified as contributing to these disparities, including bias and discrimination, lack of mentorship and networking opportunities, and lack of access to resources and support (Crown et al., 2021). Thus, these factors must be accounted for when investigating the impact of program components on career trajectories.

Appendix A: Methodology: Literature Review

A rapid literature review was conducted to identify relevant published literature on best practices for defining and tracking trainee outcomes in biomedical and genomics research (**Figure 3**).

Research Questions: The following research questions guided our literature review for the toolkit and this supplemental report:

1. What is considered successful in each trainee program?
2. What are best practices for defining and tracking trainee outcomes?
3. What are challenges and barriers to tracking trainee career development?
4. How have programs addressed the challenges and barriers to tracking trainee career development?

Guided by these research questions, we conducted a search for scientific peer-reviewed literature and technical and evaluation reports from nonprofits, professional societies (genetics, higher education), and state or federal government agencies published in the last 10 years. Searches were done with a combination of following key terms including but not limited to:

- "Longitudinal tracking," "summer research," "genetics," "genomics," "career outcomes," "biomedical"

- "Career satisfaction," "research self-efficacy," "biomedical training program"

- "Program evaluation," "trainee evaluation," "impact evaluation," "progress tracking," "trainee motivation," "engagement measurement," "program outcomes," "outcome evaluation"

Official guidelines for PRISMA flowchart for conducting literature review were followed (Page et al., 2021). Literature was identified from databases such as Google scholar (n=48), PubMed (n=23), Scopus (n=13), and additional articles were identified through sources like white papers and dissertations (n=17), and references within the peer-reviewed articles (n=17). After removing duplicates (n=37), 75 articles were screened using the inclusion and exclusion criteria agreed upon by the research team.

Inclusion criteria:

- Timeframe: Primarily focus on resources from the last five years

- Types of programs: Genetics, Biomedical, Clinical, Bioethics

Exclusion criteria:

- **Career stages:** High school graduates to early career investigators in both academic and non-academic settings.

- **Geography:** Non-U.S. studies

- **Other criteria:** Studies published more than 10 years ago, and studies focusing on social science or fine arts training programs were excluded from the review.
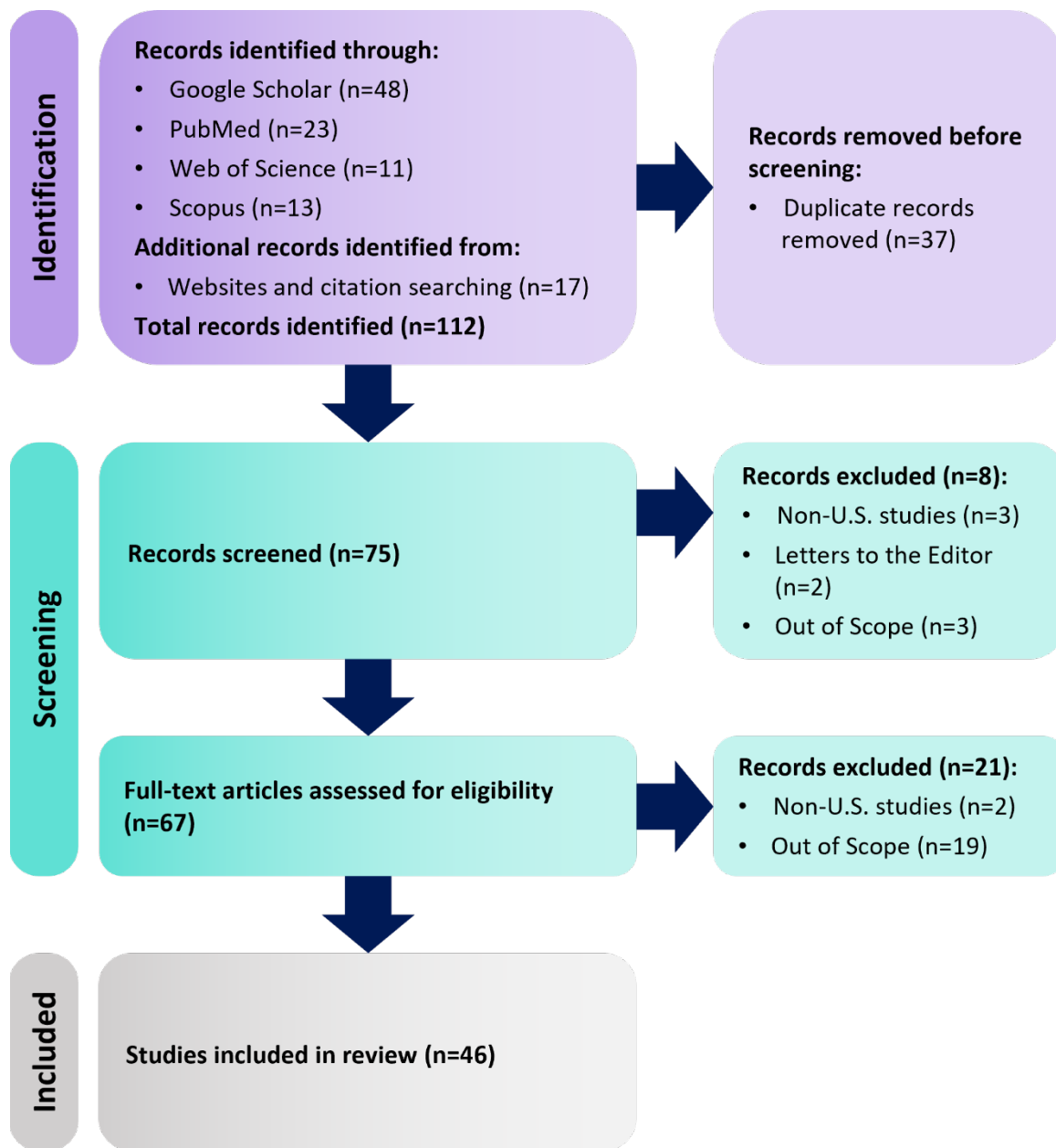
Figure 3. *PRISMA flowchart for review of literature*

Appendix B: Methodology: NIH Staff Interviews

Ripple Effect conducted five total interviews between November 1, 2022 and March 15, 2023. Interviewees were intramural and extramural program directors identified by TiDHE and were meant to have representative knowledge of NHGRI training programs. An additional NIH interviewee was identified since additional questions arose following NHGRI staff interviews. Initial outreach to request interviews with potential participants was conducted by NHGRI. Ripple Effect then followed up to schedule and conduct interviews. The interviewees included six program directors from NHGRI and the NIH Office of Extramural Research. Each interview lasted 60-75 minutes.

National Human Genome Research Institute
Advancing human health through genomics research

Ripple Effect prepared a semi-structured interview guide, which NHGRI approved following two review and feedback rounds. The interview guide included questions and follow-up probes aimed to collect information on challenges and barriers to evaluating trainee outcomes and success, different data collection methods from metrics to frequency, NHGRI's efforts for evaluating trainee success, definitions of success in trainee outcomes, and ideal trainee outcome tracking system. A second interview guide was developed to address questions that arose following the NHGRI staff interviews. These questions were organized by key topics that included barriers, defining, and measuring success, mindset and expectations, longitudinal evaluation of trainees, and networking and mentoring.

All interviews were held virtually using Microsoft Teams with one participant and one interviewer. A notetaker was also present. Five interviews were recorded with the participant's consent. Transcripts were generated using Teams live transcription for all recorded interviews. Summaries were prepared for each interview by the notetaker after thorough review of transcripts and interview notes. Two Ripple Effect researchers discussed the themes captured in each interview and deliberated until a final consensus was reached on the overall themes heard in the interviews. A summary report was generated by Ripple Effect that captured these major themes from the interviews. Interviewees were quoted without identification of names. However, participants were not considered to be anonymous as they were selected by NHGRI at the beginning of the project. Interviewees were aware of this fact and were reminded of it when they began the interview.

Appendix C: Methodology: Training Program Inventory Database

Rationale for Selecting Training Programs

A training program inventory was developed to examine characteristics of training programs across the United States to determine best practices to identify and evaluate trainee outcomes. To design the inventory, programs were first sorted into career categories of interest. These include undergraduate, graduate, postdoctoral, and genetic and clinical programs. Specific criteria were then applied to each career as follows.

- **Undergraduate programs:** Selected programs were based on age since inception, whether programs were in a highly ranked institution, and whether they were funded by a federal grant program. When possible, programs tailored to historically underrepresented populations in STEM and biomedical research were included. Summer internship programs and long-term multi-year programs were selected to include a wide range of programs.

- **Graduate programs:** Similar selection criteria described for undergraduate programs were used.

- **Postdoctoral programs:** Programs selected for this category were both federal and non-federally funded. Research was done to ensure that at least one postdoctoral program that worked with trainees that chose non-academic career paths was included.

- **Genetic counseling and clinical programs:** Only accredited programs by the Accreditation Council for Genetic Counseling (ACGC) were considered for selection of these programs. Programs with the longest history of identifying and evaluating trainees were prioritized.

After developing an initial set of programs, selected programs were reviewed to prioritize programs with a demonstrated history of evaluating trainees and to determine whether the program focused on trainees from underrepresented groups. Training programs were selected as examples with specific characteristics and are not meant to be a full representation of all training programs, or even a representative sample.

The final training program inventory included 20 training programs as shown in **Table 5**.

Table 5. *List of programs selected to develop training program inventory database*

Undergraduate Programs
- The Meyerhoff Scholarship Program   - NIGMS BUILD   - California State University, Long Beach   - California State University, Northridge   - Morgan State University   - Portland State University   - San Francisco State University   - University of Alaska, Fairbanks

- o   University of Detroit, Mercy

- o   University of Maryland, Baltimore County

- o   The University of Texas at El Paso

- o   Xavier University of Louisiana

- The Diversity Postdoctoral Alliance Committee (DPAC) HBCU Mentoring Program at Morgan State

- NINDS Health Disparities in Tribal Communities Summer Internship Program (HDTC-SIP)

- United Tribes Technical College

Graduate Programs

- Johns Hopkins Cell, Molecular, Developmental Biology, and Biophysics

- Stanford's Cellular and Molecular biology training program

- Howard University (Ph.D. Program in Biochemistry & Molecular Biology)

Postdoctoral Training/Fellowship Programs

- HHMI's Hanna H. Gray Fellows Program

- NASEM's Air Force Science & Technology Fellowship Program

- CDC's National Center for Health Statistics Postdoctoral Research Program

- The National Cancer Institute's Cancer Research Training Award Recipients

- MOSAIC Postdoctoral Career Transition Award to Promote Diversity

- FDA/ Oak Ridge Institute for Science and Education (ORISE) Postgraduate Research Program (NCTR)

- AAAS Science & Technology Policy Fellowship

- Postdoctoral Training | NYU Langone Health

Genetic Counseling and Clinical Programs

- Sarah Lawrence College

- Stanford University

- Northwestern University

- Baylor's Medical Genetics and Genomics Residency

Data Collection

Before collecting data, fields of interest were determined based on the specific research questions mentioned in **Appendix A: Methodology: Literature Review.** Fields of interest are shown in **Figure 4**. These fields captured data about each training program, successful practices to identify and evaluate trainee outcomes, and any evaluation results that were available. Initially, training program websites were used for data collection. The key in **Figure 4** demonstrates the availability of information on

program websites that was collected. For information not publicly available, program representatives including administrators and program coordinators were contacted individually by phone and through email. Contact methodology varied based on the availability of information for each program. Attempts to contact programs resulted in six total responses. Five program representatives responded to email and provided additional information whereas one responded by phone. It should be noted that of the representatives that responded to outreach, three were from undergraduate programs, one was from a graduate program, and two came from postdoctoral programs.

Training Program Inventory

A database for all selected programs from **Table 5** with the information collected in **Figure 4** made up the training program inventory. Within this database, each tab lists a career category. For each category, the selected training program is listed along with the information of interest. NIH BUILD has ten unique programs at various institutions and as a result, its own tab was created within the file to collect this information.

Training Program Database

- Program name and information; basic descriptive information in the form of an institutional profile
- Program details including its history, size, purpose, population, funding source, and other easily accessible information

- Degree to which each program tracks trainee success after program completion and the methods they use to track trainee success
- Detailed tracking information including any forms or database fields used to track trainees
- How each program defines success

- The future vision for each program, including any changes they are currently implementing to track trainee outcomes
- Any evaluations conducted to evaluate program success and details on the rigor and recency of that evaluation
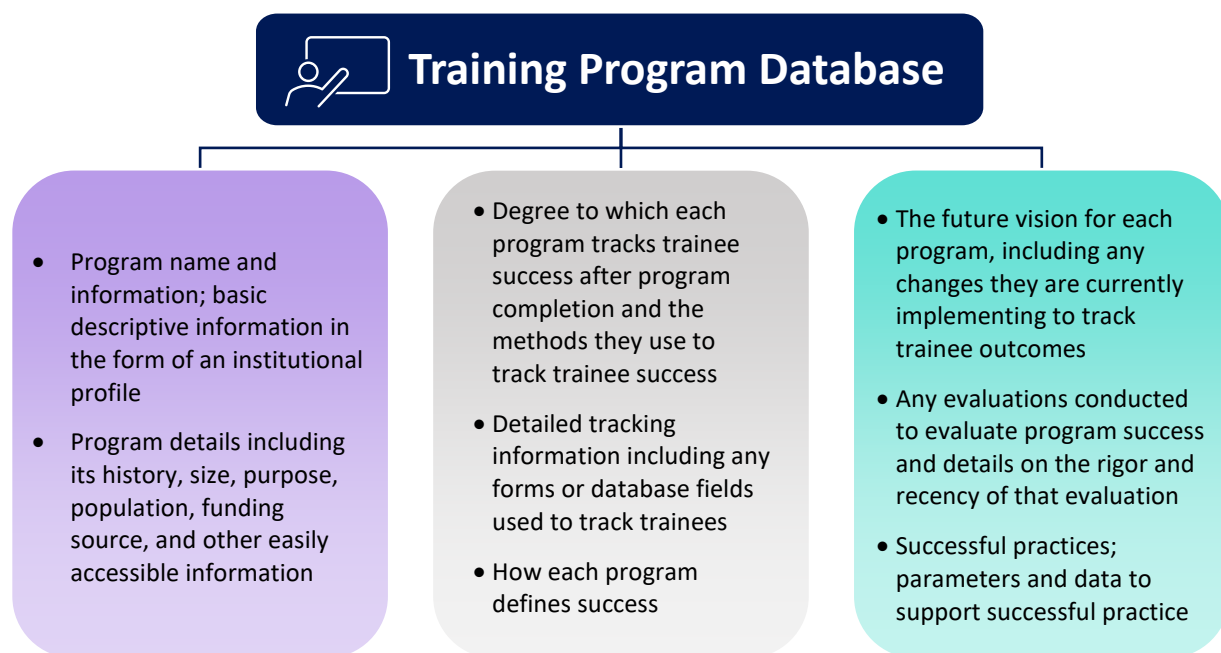- Successful practices; parameters and data to support successful practice

Figure 4. *Training program database*

Available for all programs (basic information on programs)

Available for some programs (outcomes and success definitions)

Not available for any of the 20 programs reviewed (Methods of tracking trainee success)

References

Ahmed, T., Johnson, J., Latif, Z., Kennedy, N., Javier, D., Stinson, K., & Vishwanatha, J. K. (2021).

MyNRMN: A national mentoring and networking platform to enhance connectivity and diversity

in the biomedical sciences. *FASEB BioAdvances*, *3*(7), 497–509.

https://doi.org/10.1096/fba.2020-00102

Alegría, M., Fukuda, M., Lapatin Markle, S., & NeMoyer, A. (2019). Mentoring future researchers: Advice

and considerations. *American Journal of Orthopsychiatry*, *89*, 329–336.

https://doi.org/10.1037/ort0000416

Allen-Ramdial, S.-A. A., & Campbell, A. G. (2014). Reimagining the Pipeline: Advancing STEM Diversity,

Persistence, and Success. *BioScience*, *64*(7), 612–618. https://doi.org/10.1093/biosci/biu076

Barker, E. R. (2006). Mentoring—A complex relationship. *Journal of the American Academy of Nurse

Practitioners*, *18*(2), 56–61. https://doi.org/10.1111/j.1745-7599.2006.00102.x

Brown, A. M., Chipps, T. M., Gebretsadik, T., Ware, L. B., Islam, J. Y., Finck, L. R., Barnett, J., & Hartert, T.

V. (2018). Training the next generation of physician researchers – Vanderbilt Medical Scholars

Program. *BMC Medical Education*, *18*(1), 5. https://doi.org/10.1186/s12909-017-1103-0

Byars-Winston, A., Gutierrez, B., Topp, S., & Carnes, M. (2017). Integrating Theory and Practice to

Increase Scientific Workforce Diversity: A Framework for Career Development in Graduate

Research Training. *CBE-Life Sciences Education*, *10*(4). https://doi.org/10.1187/cbe.10-12-0145

Campos, J. S., Wherry, E. J., Shin, S., & Ortiz-Carpena, J. F. (2021). Challenging systemic barriers to

promote the inclusion, recruitment, and retention of URM faculty in STEM. *Cell Host & Microbe*,

29(6), 862–866. https://doi.org/10.1016/j.chom.2021.04.001

Carini, R. M., Kuh, G. D., & Klein, S. P. (2006). Student Engagement and Student Learning: Testing the

Linkages* | SpringerLink. *Research in Higher Education*, *47*, 1–32.

https://doi.org/10.1007/s11162-005-8150-9

Carroll, M. A., Boynes, S., Jerome-Majewska, L. A., & Topp, K. S. (2022). The imperative for scientific

societies to change the face of academia: Recommendations for immediate action. *The*

Anatomical Record, *305*(4), 1019–1031. https://doi.org/10.1002/ar.24735

Crown, A., Berry, C., Khabele, D., Fayanju, O. M., Cobb, A., Backhus, L., Smith, R. N., Sweeting, R.,

Hasson, R. M., Johnson-Mann, C., Oseni, T., Newman, E. A., Turner, P., Karpeh, M., Pugh, C.,

Jordan, A. H., Henry-Tillman, R., & Joseph, K.-A. (2021). The Role of Race and Gender in the

Career Experiences of Black/African American Academic Surgeons: A Survey of the Society of

Black Academic Surgeons and a Call to Action. *Annals of Surgery*, *273*(5), 827.

https://doi.org/10.1097/SLA.0000000000004502

Cuadra, L. J., Aure, M. R. K. L., & Gonzaga, G. L. (2019). The Use of Tracer Study in Improving

Undergraduate Programs in the University. *Asia Pacific Higher Education Research Journal*

(APHERJ), *6*(1), Article 1.

https://po.pnuresearchportal.org/ejournal/index.php/apherj/article/view/1315

Dockry, M., Sachdeva, S., Fisher, C., Kenefic, L., Lacke, D., & Westphal, L. (2022). Student trainee and

paid internship programs have positive results but do little to influence long-term employee

diversity in the USDA forest service. *PLOS ONE*, *17*(11), e0277423.

https://doi.org/10.1371/journal.pone.0277423

Dugan, J., & Komives, S. (2007). Developing Leadership Capacity in College Students. In *National*

Clearinghourse for Leadership Programs.

https://www.academia.edu/download/36029451/MSLReport-FINAL.pdf

Durlak, J. A., & DuPre, E. P. (2008). Implementation Matters: A Review of Research on the Influence of

Implementation on Program Outcomes and the Factors Affecting Implementation. *American*

Journal of Community Psychology, *41*(3), 327–350. https://doi.org/10.1007/s10464-008-9165-0

Eby, L. T., Allen, T. D., Evans, S. C., Ng, T., & DuBois, D. L. (2008). Does mentoring matter? A multidisciplinary meta-analysis comparing mentored and non-mentored individuals. *Journal of Vocational Behavior*, *72*(2), 254–267. https://doi.org/10.1016/j.jvb.2007.04.005

Feldon, D. F., Litson, K., Jeong, S., Blaney, J. M., Kang, J., Miller, C., Griffin, K., & Roksa, J. (2019). Postdocs' lab engagement predicts trajectories of PhD students' skill development. *Proceedings of the National Academy of Sciences*, *116*(42), 20910–20916. https://doi.org/10.1073/pnas.1912488116

Fetterman, D. M., & Wandersman, A. (2005). *Empowerment Evaluation Principles in Practice*. Guilford Press.

Fuhrmann, C. N. (2016). Enhancing Graduate and Postdoctoral Education To Create a Sustainable Biomedical Workforce. *Human Gene Therapy*, *27*(11), 871–879. https://doi.org/10.1089/hum.2016.154

Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. J. (2016). *Impact Evaluation in Practice, Second Edition*. World Bank Publications.

Gibbs, K. D., Jr, McGready, J., Bennett, J. C., & Griffin, K. (2014). Biomedical Science Ph.D. Career Interest Patterns by Race/Ethnicity and Gender. *PLOS ONE*, *9*(12), e114736. https://doi.org/10.1371/journal.pone.0114736

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, *42*(2), 377–381. https://doi.org/10.1016/j.jbi.2008.08.010

Hatry, H. P., Newcomer, K. E., & Wholey, J. S. (2015). Evaluation Challenges, Issues, and Trends. In *Handbook of Practical Program Evaluation* (pp. 816–832). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119171386.ch31

Heck, R. H. (2004). *Studying Educational and Social Policy: Theoretical Concepts and Research Methods*.

Routledge.

Hill, L. D., Avenevoli, S., & Gordon, J. A. (2022). The Role of the National Institute of Mental Health in

Promoting Diversity in the Psychiatric Research Workforce. *Psychiatric Clinics of North America*,

45(2), 303–312. https://doi.org/10.1016/j.psc.2022.03.001

Hsu, C.-L., Liao, Y.-C., Lee, C.-W., & Chan, L. K. (2022). Privacy Concerns and Information Sharing: The

Perspective of the U-Shaped Curve. *Frontiers in Psychology*, *13*.

https://www.frontiersin.org/articles/10.3389/fpsyg.2022.771278

Hund, A. K., Churchill, A. C., Faist, A. M., Havrilla, C. A., Love Stowell, S. M., McCreery, H. F., Ng, J.,

Pinzone, C. A., & Scordato, E. S. C. (2018). Transforming mentorship in STEM by training

scientists to be better leaders. *Ecology and Evolution*, *8*(20), 9962–9974.

https://doi.org/10.1002/ece3.4527

Ilonze, O. J., Avorgbedor, F., Diallo, A., & Boutjdir, M. (2022). Addressing challenges faced by

underrepresented biomedical investigators and efforts to address them: An NHLBI-PRIDE

perspective. *Journal of the National Medical Association*, *114*(6), 569–577.

https://doi.org/10.1016/j.jnma.2022.09.007

Javier, D., Stinson, K., Zavala, M., Ahmed, T., & Vishwanatha, J. K. (2021). NRMNet: Building a National

Resource for Mentorship, Networking and Professional Development to Enhance Diversity.

Ethnicity & Disease, *31*(3), 469–480. https://doi.org/10.18865/ed.31.3.469

Jones, J., Barth, K. S., Brown, D. G., Halliday, C. A., Brady, K. T., Book, S. W., Bristol, E. J., & Back, S. E.

(2022). The Drug Abuse Research Training (DART) Program for Psychiatry Residents and Summer

Fellows: 15-Year Outcomes. *Academic Psychiatry*, *46*, 317–324. https://doi.org/10.1007/s40596-

022-01593-5

Jones, L., & Wells, K. (2007). Strategies for Academic and Clinician Engagement in Community-Participatory Partnered Research. *JAMA*, *297*(4), 407. https://doi.org/10.1001/jama.297.4.407

Jorde, L. B., Goddard, K., Aldrich, M., Chung, W., Irons, M., Jin, P., & McCague, A. (2021). *ASHG Government and Public Advocacy Committee*.

Khandker, S. R., Koolwal, G. B., & Samad, H. A. (2009). *Handbook on Impact Evaluation: Quantitative Methods and Practices*. World Bank Publications.

Langin, K. (2019). In a first, U.S. private sector employs nearly as many Ph.D.s as schools do. *Science*. https://www.science.org/content/article/first-us-private-sector-employs-nearly-many-phds-schools-do

Lauer, M., Tabak, L., & Collins, F. (2017). The Next Generation Researchers Initiative at NIH. *Proceedings of the National Academy of Sciences*, *114*(45), 11801–11803. http://www.pnas.org/cgi/doi/10.1073/pnas.1716941114

Lenzi, R. N., Korn, S. J., Wallace, M., Desmond, N. L., & Labosky, P. A. (2020). The NIH "BEST" programs: Institutional programs, the program evaluation, and early data. *The FASEB Journal*, *34*(3), 3570–3582. https://doi.org/10.1096/fj.201902064

Levy, R. (2014). Postdoc mentorship can launch careers: With high stakes for jobs in science and math, mentors must provide postdocs with a recipe for success. *American Scientist*, *102*(6), 418–422. https://go.gale.com/ps/i.do?p=AONE&sw=w&issn=00030996&v=2.1&it=r&id=GALE%7CA389260646&sid=googleScholar&linkaccess=abs

Link, A. N., & Vonortas, N. S. (2013). *Handbook on the Theory and Practice of Program Evaluation*. Edward Elgar Publishing.

Lorden, J. F., Kuh, C. V., & Voytuk, J. A. (2011). *Research-Doctorate Programs in the Biomedical Sciences: Selected Findings from the NRC Assessment*.

Mathur, A., Hwalek, M., Straub, V., & Chow, C. S. (2023). Increasing faculty support, respect, and ability

    to help doctoral students explore non-academic research career opportunities. *Heliyon*, *9*(1),

    e13052. https://doi.org/10.1016/j.heliyon.2023.e13052

Mentor Collective. (2020). *Mentor Collective Programs Increase Sense of Belonging by 14%* [Case Study].

    https://www.mentorcollective.org/paper-sense-of-belonging

Meyers, F. J., Mathur, A., Fuhrmann, C. N., O'Brien, T. C., Wefes, I., Labosky, P. A., Duncan, D. S., August,

    A., Feig, A., Gould, K. L., Friedlander, M. J., Schaffer, C. B., Van Wart, A., & Chalkley, R. (2016).

    The origin and implementation of the Broadening Experiences in Scientific Training programs:

    An NIH common fund initiative. *The FASEB Journal*, *30*(2), 507–514.

    https://doi.org/10.1096/fj.15-276139

National Academies of Sciences, Engineering, and Medicine. (2018). *Graduate STEM Education for the*

    21st Century. https://nap.nationalacademies.org/catalog/25038/graduate-stem-education-for-

    the-21st-century

Newcomer, K. E., Hatry, H. P., & Wholey, J. S. (2015). Practical Data Collection Procedures. In *Handbook*

    of Practical Program Evaluation (pp. 319–324). John Wiley & Sons, Ltd.

    https://doi.org/10.1002/9781119171386.part2

NIH. (2012). *National Institutes of Health: Draft Report of the Advisory Committee to the Director*

    Working Group on Diversity in the Biomedical Research Workforce. NIH.

    https://acd.od.nih.gov/documents/reports/DiversityBiomedicalResearchWorkforceReport.pdf

NIH. (2017). *NIH Office of the Senior Official for Privacy*.

    https://oma.od.nih.gov/DMS/Documents/Privacy/NIH%20Privacy%20FAQs%202017.pdf

NSF. (2020). *Doctorate Recipients from U.S. Universities: 2020 | NSF - National Science Foundation*.

    https://ncses.nsf.gov/pubs/nsf22300/report

Odedina, F. T., Behar-Horenstein, L. S., Fathi, P., Kaninjing, E., Nguyen, J., Askins, N., Reams, R. R.,

Mochona, B., Ezeani, A., Lyon, D., Jacobs, K., & Maduka, M. (2022). Improving Representation of

Underrepresented Minority (URM) Students in Oncology Biomedical Research Workforce:

Outcome Evaluation from the ReTOOL Program. *Journal of Cancer Education*, *37*(1), 37–45.

https://doi.org/10.1007/s13187-020-01779-1

Ognibene, F. P., Gallin, J. I., Baum, B. J., Wyatt, R. G., & Gottesman, M. M. (2016). Outcomes From the

NIH Clinical Research Training Program: A Mentored Research Experience to Enhance Career

Development of Clinician–Scientists. *Academic Medicine*, *91*(12), 1684.

https://doi.org/10.1097/ACM.0000000000001245

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L.,

Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A.,

Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA

2020 statement: An updated guideline for reporting systematic reviews. *BMJ (Clinical Research

Ed.)*, *372*, n71. https://doi.org/10.1136/bmj.n71

Patton, M. Q. (2008). *Utilization-Focused Evaluation*. SAGE Publications.

Petrie, K. A., Carnahan, R. H., Brown, A. M., & Gould, K. L. (2017). Providing Experiential Business and

Management Training for Biomedical Research Trainees. *CBE Life Sciences Education*, *16*(3),

ar51. https://doi.org/10.1187/cbe.17-05-0074

Pike, G. R., Kuh, G. D., & McCormick, A. C. (2011). An Investigation of the Contingent Relationships

Between Learning Community Participation and Student Engagement. *Research in Higher

Education*, *52*(3), 300–322. https://doi.org/10.1007/s11162-010-9192-1

Poister, T. H. (2015). Performance Measurement. In *Handbook of Practical Program Evaluation* (pp. 108–

136). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119171386.ch5

Polka, J. K., Krukenberg, K. A., & McDowell, G. S. (2015). A call for transparency in tracking student and

postdoc career outcomes. *Molecular Biology of the Cell*, *26*(8), 1413–1415.

https://doi.org/10.1091/mbc.E14-10-1432

Ravenscroft, J., Liakata, M., Clare, A., & Duma, D. (2017). Measuring scientific impact beyond academia:

An assessment of existing impact metrics and proposed improvements. *PLOS ONE*, *12*(3),

e0173152. https://doi.org/10.1371/journal.pone.0173152

Rice, T. K., Jeffe, D. B., Boyington, J. E. A., Jobe, J. B., Dávila-Román, V. G., Gonzalez, J. E., De Las Fuentes,

L., Makala, L. H. C., Sarkar, R., Ogedegbe, G. G., Taylor, A. L., Czajkowski, S., Rao, D. C., Pace, B.

S., Jean-Louis, G., & Boutjdir, M. (2017). Mentored training to increase diversity among faculty in

the biomedical sciences: The NHL BI Summer Institute Programs to Increase Diversity (SI PID)

and the Programs to Increase Diversity among Individuals Engaged in Health-related Research

(PRIDE). *Ethnicity and Disease*, *27*(3), 249–256. https://doi.org/10.18865/ed.27.3.249

Rice, T. K., Jeffe, D. B., Jobe, J. B., Boutjdir, M., Pace, B. S., & Rao, D. C. (2014). Enhancing the Careers of

Under-Represented Junior Faculty in Biomedical Research: The Summer Institute Program to

Increase Diversity (SIPID). *Journal of the National Medical Association*, *106*(1), 50–57.

https://doi.org/10.1016/s0027-9684(15)30070-5.

Rossi, P. H., Lipsey, M. W., & Henry, G. T. (2018). *Evaluation: A Systematic Approach*. SAGE Publications.

Saetermoe, C. L., Chavira, G., Khachikian, C. S., Boyns, D., & Cabello, B. (2017). Critical race theory as a

bridge in science training: The California State University, Northridge BUILD PODER program.

BMC Proceedings, *11*(12), 21. https://doi.org/10.1186/s12919-017-0089-2

Sauermann, H., & Roach, M. (2016). Why pursue the postdoc path? | Science. *Science*, *352*(6286), 663–

664. https://www.science.org/doi/abs/10.1126/science.aaf2061

Scalo, J., & Freauff, L. (2020). *Evaluation of the National Institutes of Health (NIH) Broadening*

   Experiences in Scientific Training (BEST) Program.

   https://commonfund.nih.gov/sites/default/files/NIH_BEST_Program_FInal_Report_508.pdf

Schaller, J., & Stalsburg, B. (2022). *The Simple Truth about the Gender Pay Gap*.

   https://www.aauw.org/app/uploads/2022/12/SimpleTruth_12.22_2.1-002.pdf

Segarra, V. A., Blatch, S., Boyce, M., Carrero-Martinez, F., Aguilera, R. J., Leibowitz, M. J., Zavala, M.,

   Hammonds-Odie, L., & Edwards, A. (2020). Scientific Societies Advancing STEM Workforce

   Diversity: Lessons and Outcomes from the Minorities Affairs Committee of the American Society

   for Cell Biology. *Journal of Microbiology & Biology Education*, *21*(1), 15.

   https://doi.org/10.1128/jmbe.v21i1.1941

Sorkness, C. A., Pfund, C., Ofili, E. O., Okuyemi, K. S., Vishwanatha, J. K., Zavala, M. E., Pesavento, T.,

   Fernandez, M., Tissera, A., Deveci, A., Javier, D., Short, A., Cooper, P., Jones, H., Manson, S.,

   Buchwald, D., Eide, K., Gouldy, A., Kelly, E., … on behalf of the NRMN team. (2017). A new

   approach to mentoring for research careers: The National Research Mentoring Network. *BMC*

   Proceedings, *11*(12), 22. https://doi.org/10.1186/s12919-017-0083-8

Spence, J. P., Buddenbaum, J. L., Bice, P. J., Welch, J. L., & Carroll, A. E. (2018). Independent investigator

   incubator (I3): A comprehensive mentorship program to jumpstart productive research careers

   for junior faculty. *BMC Medical Education*, *18*(1), 186. https://doi.org/10.1186/s12909-018-

   1290-3

Springer Nature. (2019). *Measures of success: What's the real value of published research?*

   https://www.springernature.com/gp/librarians/news-events/all-news-

   articles/journals/measures-of-success/16494450

Straus, S. E., Johnson, M. O., Marquez, C., & Feldman, M. D. (2013). Characteristics of Successful and

   Failed Mentoring Relationships: A Qualitative Study Across Two Academic Health Centers.

National Human Genome Research Institute
Advancing human health through genomics research

Academic Medicine : Journal of the Association of American Medical Colleges, *88*(1), 82–89.

https://doi.org/10.1097/ACM.0b013e31827647a0

Thompson, R. C., Monroe-White, T., Xavier, J., Howell, C., Moore, M. R., & Haynes, J. K. (2016).

Preparation of underrepresented males for scientific careers: A study of the Dr. John H. Hopps

Jr. defense research scholars program at morehouse college. *CBE Life Sciences Education*, *15*(3),

Article 3. https://doi.org/10.1187/cbe.15-12-0263

UMBC. (2022). *Results*. https://meyerhoff.umbc.edu/about/results/

Webb, T. J., Guerau-de-Arellano, M., Jones, H. P., Butts, C. L., Sanchez-Perez, L., & Montaner, L. J. (2022).

The Minority Scientists' Experience: Challenging and Overcoming Barriers to Enhancing Diversity

and Career Advancement. *Journal of Immunology (Baltimore, Md.: 1950)*, *208*(2), 197–202.

https://doi.org/10.4049/jimmunol.2101077

Weiss, C. H., & Weiss, C. H. (1998). *Evaluation: Methods for studying programs and policies* (2nd ed).

Prentice Hall.

Wilson, A. E., Pollock, J. L., Billick, I., Domingo, C., Fernandez-Figuero, E. G., Nagy, E. S., Steury, T. D., &

Summers, A. (2018). Assessing Science Training Programs: Structured Undergraduate Research

Programs Make a Difference. *BioScience*, *68*(7), 529–534. https://doi.org/10.1093/biosci/biy052

Wright-Mair, R. (2020). Longing to Belong: Mentoring Relationships as a Pathway to Fostering a Sense of

Belonging for Racially Minoritized Faculty at Predominantly White Institutions. *Journal*

Committed to Social Change on Race and Ethnicity (JCSCORE), *6*(2), 2–31.

https://www.jstor.org/stable/48644535

Yomtov, D. (2013). *An evaluation of the CSUN building connections for success peer mentoring program*

[Thesis, California State University, Northridge].

http://dspace.calstate.edu/handle/10211.2/3256

Zambrana, R. E., Carvajal, D., & Townsend, J. (2023). Institutional penalty: Mentoring, service, perceived

    discrimination and its impacts on the health and academic careers of Latino faculty. *Ethnic and*

    Racial Studies, *0*(0), 1–26. https://doi.org/10.1080/01419870.2022.2160651

Zuniga, N. C., & Colbern, A. (2021). The Annual BUILD Snapshot: Tracking Alumni Outcomes. *HCI*

    International 2021 - Late Breaking Papers: Cognition, Inclusion, Learning, and Culture, *13096*,

    345–360. https://doi.org/10.1007/978-3-030-90328-2_22