

FINAL REPORT

Feasibility Study to Design an Outcome Evaluation of the Trans-NIH Summer Research Training Institutes Program

CONTRACT# HHSN276201300277U

Submitted by:

The Madrillon Group Inc. 8142 Madrillon Court Vienna, Virginia 22182

Submitted to:

G. Stephane Philogene, Ph.D. Jaclyn M. Crouch, M.P.A. Co-Contracting Officer Representatives National Institutes of Health Office of Behavioral and Social Sciences Research 31 Center Drive Room B1C19 Bethesda, Maryland 20892

September 2014

Table of Contents

Lis	at of Exhib	its	4
Ex	ecutive Su	Immary	6
	Introductio	วท	6
	Design ar	d Methods of the Feasibility Study	6
	Findings .		7
	Recomme	ndations	9
I.	Introduc	tion	11
	1.1 Pur	pose of the Feasibility Study	11
	1.2 Use	e of the Feasibility Study Results	11
	1.3 Ove	erview of the Feasibility Study Report	12
II.	Feasibili	ty Study Questions, Methods, and Data Sources	13
	2.1 Fea	sibility Study Evaluation Questions	13
	2.2 Dat	a Sources and Data Collection Approaches	13
	2.2.1	Document Review	15
	2.2.2	Literature Review	15
	2.2.3	Interviews with Federal Stakeholders and Training Institute Faculty	16
	2.2.4	Pilot Studies	17
III.	The Trar	s-NIH Summer Research Training Institutes	19
III.	The Tra r 3.1 His	s-NIH Summer Research Training Institutes tory of the Trans-NIH Summer Research Training Institutes Program	19 19
III.	The Tran 3.1 His 3.2 His	s-NIH Summer Research Training Institutes tory of the Trans-NIH Summer Research Training Institutes Program tory of the Seven Summer Research Training Institutes	19 19 21
III.	The Tran 3.1 His 3.2 His 3.2.1	tory of the Trans-NIH Summer Research Training Institutes Institutes Program	19 19 21 21
III.	The Tran 3.1 His 3.2 His 3.2.1 3.2.2	s-NIH Summer Research Training Institutes tory of the Trans-NIH Summer Research Training Institutes Program tory of the Seven Summer Research Training Institutes Randomized Clinical Trials Involving Behavioral Interventions (RCT) NIH Summer Institute on Social and Behavioral Intervention Research (SW).	19 19 21 21 22
ш.	The Tran 3.1 His 3.2 His 3.2.1 3.2.2 3.2.3	s-NIH Summer Research Training Institutes tory of the Trans-NIH Summer Research Training Institutes Program tory of the Seven Summer Research Training Institutes Randomized Clinical Trials Involving Behavioral Interventions (RCT) NIH Summer Institute on Social and Behavioral Intervention Research (SW). The Advanced Training Institute on Health Behavior Theory (ATI)	19 21 21 22 22
Ш.	The Tran 3.1 His 3.2 His 3.2.1 3.2.2 3.2.3 3.2.3 3.2.4	s-NIH Summer Research Training Institutes tory of the Trans-NIH Summer Research Training Institutes Program tory of the Seven Summer Research Training Institutes Randomized Clinical Trials Involving Behavioral Interventions (RCT) NIH Summer Institute on Social and Behavioral Intervention Research (SW). The Advanced Training Institute on Health Behavior Theory (ATI) Summer Institute in Applied Research on Child and Adolescent Development (C&AD)	19 21 21 22 22 22
ш.	The Trar 3.1 His 3.2 His 3.2.1 3.2.2 3.2.3 3.2.4 3.2.5	As-NIH Summer Research Training Institutes	19 21 21 22 22 22
III.	The Tran 3.1 His 3.2 His 3.2.1 3.2.2 3.2.3 3.2.4 3.2.5 3.2.6	As-NIH Summer Research Training Institutes	19 21 21 22 22 22 23 24 25
III.	The Tran 3.1 His 3.2 His 3.2.1 3.2.2 3.2.3 3.2.4 3.2.5 3.2.6 3.2.7	As-NIH Summer Research Training Institutes	19 21 21 22 22 23 24 25
Ш.	The Tran 3.1 His 3.2 His 3.2.1 3.2.2 3.2.3 3.2.4 3.2.5 3.2.6 3.2.7 3.3 Key Doo	As-NIH Summer Research Training Institutes	19 21 21 22 22 23 24 25 25 25
III. IV.	The Tran 3.1 His 3.2 His 3.2.1 3.2.2 3.2.3 3.2.4 3.2.5 3.2.6 3.2.7 3.3 Key Doo	As-NIH Summer Research Training Institutes	19 21 21 22 22 23 24 25 25 25 26 29
III. IV.	The Tran 3.1 His 3.2 His 3.2.1 3.2.1 3.2.2 3.2.3 3.2.3 3.2.4 3.2.5 3.2.6 3.2.7 3.3 Key Door Structur 4.1 Pla	As-NIH Summer Research Training Institutes	19 21 21 22 22 23 23 24 25 25 25 26 29 29
III. IV.	The Tran 3.1 His 3.2 His 3.2.1 3.2.2 3.2.3 3.2.4 3.2.5 3.2.6 3.2.7 3.3 Key Doo Structur 4.1 Pla 4.1.1	As-NIH Summer Research Training Institutes	19 21 21 22 22 23 23 24 25 25 26 26 29 29
III. IV.	The Tran 3.1 His 3.2 His 3.2.1 3.2.2 3.2.3 3.2.4 3.2.5 3.2.6 3.2.7 3.3 Key Doo Structur 4.1 Pla 4.1.1 4.1.2	As-NIH Summer Research Training Institutes	19 21 21 22 22 23 23 24 25 25 25 26 29 29 29 31

	4.1.4	Blendin	g of Instructional Methods	34
	4.1.5	The Va	lue of a Residential Immersion Learning Experience versus Online	
		Training	J	34
	4.2 Re	ecruiting a	nd Selecting Faculty, Host Institutions, and Trainees	36
	4.2.1	Recruit	ment and Selection of Faculty	36
	4.2.2	Selection	on of Host Institutions	37
	4.2.3	Recruit	ment and Selection of Trainees	38
		4.2.3.1	Target Populations	38
		4.2.3.2	Trainee Selection Process	39
		4.2.3.3	Availability and Quality of Data on Past Years' Trainees and Unselec Applicants	ted 40
		4.2.3.4	Institutional Post-Training Workplace Supports	41
	4.3 Ke	ey Finding	s on Structure and Implementation	42
v	Outcon	nes from	the Summer Research Training Institutes	45
۷.				45
	5.1 If		Models and Frameworks in the Training Evaluation Literature	45
	5.1.1	I ne Kir		45
	5.1.2	Alterna	tives to the Kirkpatrick Four Levels® Model	47
	5.2 Re In	stitutes	arlier Evaluations of NIH-Sponsored Summer Research Training	48
	5.2.1	Charac	teristics of the Summer Research Training Institutes	49
	5.2.2	Charac	teristics of Summer Research Training Institute Evaluation Designs	51
	5.3 Or Tr	utcomes le aining Ins	dentified by NIH Program Officers and Trans-NIH Summer Research titute Faculty Members	52
	5.4 Pi	lot Study of	of Trainee Satisfaction with the ISSH Training Institute	54
	5.5 Pe	erspective	s on the Use of Comparison Groups	55
	5.5.1	Perspe	ctives of NIH Program Officers and Training Institute Faculty Members	355
	5.5.2	Pilot St	udy Using Trainees and Unselected Applicants	59
	5.6 Ke	ey Finding	s on Outcomes	59
۱ <u>л</u>	Conclu	-!		C O
VI.	Conciu	sions and	Recommendations	62
	6.1 Fe	easibility o	f an Outcome Evaluation	62
	6.2 Re	ecommen	ded Design for the Summer Research Training Institutes Outcome	62
	E\ 6.2.1	Compo	nonto of the Decign	03 62
	0.2.1	6 2 1 1	Multiple Case Studies	03
		0.2.1.1	Success Case Studies	05
	600	0.2.1.2	rison Stratogics Used in the Outcome Evolution Decise	00
	0.2.2	Drimor	rison Suralegies Used in the Outcome Evaluation Design	00 67
	0.2.3		and Secondary Outcome Evaluation Questions	07 70
	0.2.4		ie measures, methos, and Data Sources	0 ז כד
	0.2.3	6251	NIL Databases and Internal Decords	ני כד
		0.2.3.1 6 3 5 3	NIII Databases and internal Records	13 71
		0.2.5.2		/4

Appendix		7	'8
References .		7	'6
	6.2.5.4	Success Case Study Interviews7	4
	6.2.5.3	Trainee Online Survey7	4

List of Exhibits

Exhibit 1. SRTI Outcome Evaluation Feasibility Study Questions	
Exhibit 2. Methodology for the SRTI Outcome Evaluation Feasibility Study	
Exhibit 3. Feasibility Study Questions, Data Sources and Data Collection Approaches14	
Exhibit 4. Document and Administrative Records Included in the Document Review15	
Exhibit 5. Literature Review Questions16	
Exhibit 6. The Seven Summer Research Training Institutes	
Exhibit 7. RCT Core Curriculum Topics21	
Exhibit 8. Conceptual Model for the Trans-NIH Summer Research Training Institutes Program28	
Exhibit 9. Curriculum Content	
Exhibit 10. Changes to Curriculum	
Exhibit 11. Post-Training Support Mechanisms	
Exhibit 12. Individual Projects	
Exhibit 13. Host Institutions	
Exhibit 14. Target Population by Career Stage	
Exhibit 15. Other Relevant Target Population Variables/Considerations	
Exhibit 16. Participant Selection Materials	
Exhibit 17. The Kirkpatrick Four Levels® Model as a Planning and Evaluation Approach46	
Exhibit 18. Summer Research Training Institutes, Sponsoring Institutes, and Evaluation Study Citations	
Exhibit 19. Outcome Variables Used In Summer Research Training Institute Evaluations51	
Exhibit 20. Evaluation Designs Used in Summer Research Training Institute Evaluations51	
Exhibit 21. Outcome Categories Identified by NIH Program Officers and SRTI Faculty53	
Exhibit 22. NIH Program Officers' Suggested Comparison Groups	
Exhibit 23. Outcome Chain Based on Interviewee Responses	
Exhibit 24. Logic Model for the Trans-NIH Summer Research Training Institutes Outcome Evaluation	
Exhibit 25. Primary and Secondary Outcome Evaluation Questions and Proposed Comparisons68	3
Exhibit 26. Key Variables, Metrics, and Data Sources70	
Exhibit 27. Outcome Evaluation Data Sources	

Acknowledgements

The authors of this report thank the many people who contributed to this evaluation and to the conceptualization, design, and conduct of the *Feasibility Study to Design an Outcome Evaluation of the Trans-NIH Summer Research Training Institutes Program.* The evaluation could not have been completed without the cooperation and responsiveness of staff members from the Office of Behavioral and Social Sciences Research, especially Dr. G. Stephane Philogene and Jaclyn M. Crouch, who were the co-Contracting Officer Representatives for the project. We also thank the following individuals who served as members of the Evaluation Advisory Committee: Drs. Sarah Kobrin (National Cancer Institute), Peter Kaufmann (National Heart, Lung, and Blood Institute), Patty Mabry (Office of Disease Prevention), Valerie Maholmes (Eunice Kennedy Shriver National Institute of Child Health and Human Development), Helen Meissner (Office of Disease Prevention), Wendy Nilsen (Office of Behavioral and Social Sciences Research), and G. Stephane Philogene (Office of Behavioral and Social Sciences Research). Each of these individuals provided the authors with guidance, thoughtful feedback, and support throughout the project.

Executive Summary

Introduction

One of the primary goals of the Office of Behavioral and Social Sciences Research (OBSSR) is to foster and promote a behavioral and social sciences interdisciplinary perspective within the research portfolios of the National Institutes of Health (NIH). The integration of behavioral and social sciences theories and methods within biomedical research requires research investigators to possess specific knowledge and skills that encompass both of these broad domains. When OBSSR was established in 1999, program staff members across NIH had recognized that research proposals bridging these domains were not faring well in the internal grant peer review process. There was a clear need to familiarize behavioral and social scientists and biomedical researchers with basic components of each other's fields.

Beginning in FY 2001, OBSSR and the National Heart, Lung, and Blood Institute (NHLBI) jointly sponsored the first summer research training institute on Randomized Clinical Trials Involving Behavioral Interventions. This nine-day intensive residential educational experience proved highly popular and was conducted on an annual basis thereafter. The summer research training institute model was replicated through a series of partnerships between OBSSR, other NIH Institutes and Offices, and some outside federal and non-federal agencies and organizations. Over the following ten years, the seven trans-NIH summer research training institutes that grew from this initial offering came to form the trans-NIH Summer Research Training Institute Program. The absence of a broader outcome evaluation across the seven training institutes, and growing administrative concerns about the justification for continued financial underwriting of the costs of providing these training opportunities, led OBSSR to commission an Evaluation Feasibility Study to determine how best to design an outcome evaluation study for the seven training institutes. On September 20, 2013, OBSSR commissioned the Madrillon Group Inc. to conduct a Feasibility Study to Design an Outcome Evaluation of the Trans-NIH Summer Research Training Institutes Program (Feasibility Study)¹. This report presents the design, methods, results and recommendations from that study.

Design and Methods of the Feasibility Study

The Feasibility Study was designed as a mixed-methods cross-sectional study in order to answer several questions concerning whether an outcome evaluation of the seven training institutes was feasible and warranted, what types of outcome questions and outcomes could be examined in an outcome evaluation, what types of indicators and measures would be appropriate, whether any type of comparison group was feasible, and what type of evaluation design would be most appropriate. The Feasibility Study utilized several sources of archival data and new data collected specifically for the study. Archival sources included a review of web pages established for several training institutes, records of trainees and unselected applicants, and satisfaction questionnaires provided for four annual sessions of one training institute. The Madrillon evaluation team conducted a literature review of 18 published evaluations of other

¹ Formal work on the project began after October 15, 2013 as a result of the Federal government shutdown.

summer research training institutes at NIH during the past fifteen years. The evaluation team also collected new data in the form of interviews with 16 NIH program officers who had been active in planning and organizing the seven training institutes, and 6 outside (non-Federal) faculty members who taught at the training institutes.

The Madrillon evaluation team conducted two pilot studies as part of the overall Evaluation Feasibility Study. The first was an analysis of research grant and publication productivity during the two years following the 2011 session of the *Training Institute for Dissemination and Implementation Research in Health*, using 33 participating trainees, and a random sample of unselected applicants for that session. The second pilot study was an analysis of 200 participating trainees' training satisfaction questionnaires from four sessions of the *Institute on Systems Science and Health*.

Findings

The Evaluation Feasibility Study produced the following general findings.

Orientation and Structure of the Seven NIH Summer Research Training Institutes

- The seven summer research training institutes address different scientific fields and methodological issues. Three training institutes (*Randomized Clinical Trials with Behavioral Interventions, Summer Institute in Applied Research in Child and Adolescent Development*, and *Advanced Training Institute on Health Behavior Theory*) addressed scientific fields and methodologies that are scientifically well established. The fourth training institute (*NIH Summer Institute on Social and Behavioral Intervention Research*) sought to build research capacity among doctoral-level social workers. The last three training institutes (*Institute on Systems Science and Health, Training Institute on Dissemination and Implementation Research in Health, and the m*(obile) *Health Training Institute*) addressed three rapidly emerging scientific fields of inquiry.
- The three training institutes on emerging fields were the most recently launched and involved outside partners in addition to NIH Institutes and Offices. The *Summer Institute in Applied Research in Child and Adolescent Development* also involved a non-NIH partner.
- Two of the emerging fields training institutes (*Institute on Systems Science and Health* and *Training Institute on Dissemination and Implementation Research in Health*) and the social work training institute were embedded in broader initiatives within NIH that included specific research funding opportunity announcements.
- The educational design of each of the seven summer research training institutes incorporated principles of adult learning theory. Adult learning theory characterizes adult learners as internally motivated individuals who learn best when training content is perceived as relevant to their personal or professional needs and is presented using a mix of instructional methods that emphasize their active participation in the learning process. When these conditions are met, adult learning theory posits that learners are more fully engaged and involved in the learning experience and therefore more likely to absorb new knowledge and learn new skills.

- The curricula developed for each of the seven training institutes differed in terms of educational content, but shared common emphasis on basic theories and methods. Four training institutes included material on grantsmanship and three included instruction on developing multidisciplinary research teams and building new research fields.
- Four components of the training institute structure were seen as critical for success: mentoring, networking with other trainees and training faculty, individual or group projects as a means of immediately applying what was taught, and the residential nature of the training experience.
- The summer research training institutes targeted potential trainees largely by career stage. The three established-field training institutes and the social work institute targeted early stage trainees (within ten years of the highest research degree), while the emerging field institutes sought both early and late stage trainees from the biomedical fields and the specific emerging fields.
- The trainee selection process was similar in all seven training institutes, but the availability of data documenting the process varied considerably. Each applicant submitted an admission package (curriculum vita, personal statement describing professional and research goals, letters of recommendation) that was rated by an institute selection committee on a set of pre-established rating criteria. A general threshold score was established, and, for the most part, trainees who scored above this threshold were accepted and those scoring below were not. Data on these scores were not uniformly and consistently available for all of the training institutes for all years.² This lack of data availability affected the evaluation team's recommendations for design options for the outcome evaluation.

Review of Published Evaluations of Other NIH Summer Research Training Institutes

- Using a carefully constructed definition of "NIH summer research training institute" programs the Madrillon Evaluation team conducted a series of bibliographic searches on PubMed, Scopus, Google, and Google Scholar. These searches yielded a total of 18 publications describing a total of 21 evaluations of previous NIH training institutes that were similar in length and focus to the seven summer research training institutes. The searches also identified 18 review articles, discussions, handbooks, and reports characterizing the current state of training evaluation practice.
- The Kirkpatrick Four-Levels[®] Model (now rebranded as the *New World Kirkpatrick Four-Levels Model*[®]) is the dominant conceptual framework that guides most evaluation practice in the training field today. The model has been adapted by the Centers for Disease Control and Prevention (Training Effectiveness and Efficiency Model--Dunet and Reyes, 2006) and the United States Office of Personnel Management's (2011) *Training and Field Guide-Demonstrating the Value of Training at Every Level.*

² After submission of the draft version of this report in August, some members of the Evaluation Advisory Committee indicated that they thought some additional data for some training institutes might exist that had not been located during the Evaluation Feasibility Study.

- Evaluations of NIH-sponsored summer research training institutes are relatively few in number but have shown a marked increase during the past five years.
- Published evaluations of NIH-sponsored summer research training institutes share several important design weaknesses, including the absence of conceptual frameworks, limited use of comparison groups, failure to identify and measure institutional factors that either supported or impeded trainees' application of what they learned, and an absence of long-term follow-up. The evaluations do not link to each other and do not seek to build a collective body of knowledge about what works and what does not work in this type of training.
- These evaluations have identified a useful set of outcomes that can be used in future studies, including subsequent research grants and research publications, career advancement, increased motivation to write grants, maintaining contact with other trainees and mentors, and use of training content in their teaching activities.

Outcomes and Suggestions for Design of an Outcome Evaluation

- Interviewees identified a series of outcomes highly similar to those identified through the literature review. They also noted that Kirkpatrick Level 4 results would apply to several groups, including universities (which might establish training programs of their own), NIH (which would experience growth in its research portfolios and issue new Funding Opportunity Announcements), and the scientific field (science advances and discoveries).
- Interviewees acknowledged that a randomized controlled experiment would provide the strongest form of evidence for training institute efficacy, but it was not practical at present due to the need to generate information within a relatively short timeframe (two years), and issues arising from randomly assigning applicants to receive the training in place of the current selection process.
- Interviewees outlined a variety of potential quasi-experimental comparison strategies, but were critical of each of them. Of those suggested, the two that elicited the most comment included comparisons of trainees' research productivity before and after the training institute session, and comparisons with unselected applicants. When possible, interviewees suggested that using scoring data on unselected applicants to generate a pool of individuals that were most comparable to those selected would be advisable. However, examination of the data provided during the Feasibility Study showed that this was possible for only a limited number of training institutes and years.
- In subsequent review of the draft version of this report, some Evaluation Advisory Committee members indicated that there might be additional data on scoring that had not been made available to the evaluation team. The extent of these data is not currently known.

Recommendations

 The New World Kirkpatrick Evaluation Model[®] is recommended to serve as the basis for the conceptual model for an Outcome Evaluation of the NIH Summer Research Training Institutes Program.

- Based on interviews with federal program officers and external faculty members involved with the seven training institutes, the Madrillon evaluation team recommends that the overall design for the outcome study should use a multiple case study approach using both quantitative and qualitative data.
- The interviewees as a group agreed that a prospective randomized experimental design would provide the strongest evidence of the effectiveness of the training institutes, but this was not feasible for the proposed Outcome Evaluation.
- The interview data did not identify a clear retrospective comparison strategy that was favored by a majority of the interviewees. The quasi-experimental comparison strategies discussed by interviewees were found to have various weaknesses which rendered them less than ideal. Despite these limitations, the two strategies considered most appropriate included pre-post comparisons of the effects of the training on trainees, and the use of a random sample of unselected applicants. There was some interest in the possible use of trainees from one training institute as comparators for trainees from other training institutes for established scientific fields with those from emerging fields as the research funding opportunities for the latter might differ from those for the former.

I. Introduction

Under a contract awarded on September 20, 2013 by the Office of Behavioral and Social Sciences Research (OBSSR) of the National Institutes of Health (NIH), the Madrillon Group Inc. conducted an evaluation feasibility study to determine whether an outcome evaluation of the trans-NIH Summer Research Training Institutes (SRTI) program is feasible and warranted, and to provide recommendations for its design. This report describes the methods and data sources for the Feasibility Study and presents the key findings. The report concludes with a series of recommendations concerning appropriate evaluation questions and design for the outcome evaluation.

1.1 Purpose of the Feasibility Study

An evaluation feasibility study typically answers the following three questions:

- 1. Is an evaluation of a program feasible and warranted at this time?
- 2. What type of evaluation approach is suitable for the program at its current level of maturity?
- 3. What evaluation design, key variables, data sources, and data collection approaches are most appropriate for the study, given the client's constraints on time, budget, and the quality and availability of data?

Mindful of evaluation recommendations published by the Institute of Medicine (Manning et al., 2004) that encourage process and outcome-oriented evaluation of major NIH research programs at least every five years, OBSSR wanted to determine the feasibility of conducting an outcome evaluation of the Summer Research Training Institutes (SRTI) program as a first step in evaluating the program. These summer training institute programs have not been evaluated beyond typical immediate post-training satisfaction questionnaires with the exception of the *Training Institute for Dissemination & Implementation Research in Health*, for which results from a six-month post-workshop evaluation were recently published (Meissner et al., 2013).

The evaluation Feasibility Study was designed to answer the above questions for the SRTI program and to identify the most appropriate evaluation questions (including identification of those questions that can be answered for all seven SRTIs, and which questions may be more appropriate for specific individual training institutes); determine an appropriate study design; identify a possible comparison group; and stipulate the key quantitative and qualitative variables, measures, data sources, and data collection approaches associated with them.

1.2 Use of the Feasibility Study Results

The findings from this study will be used to inform the full-scale evaluation of the SRTI program at NIH. The primary audiences for the Feasibility Study report include Federal policymakers within NIH and its Institutes, Centers, and Offices participating in the SRTI program. The results will be used in several ways. They will

- 1. Provide important information about the scope and influence of the NIH SRTI program;
- 2. Inform the outcome evaluation design and methods to be used;

- 3. Offer valuable insight into how to integrate a biobehavioral interdisciplinary perspective into other NIH research areas; and
- 4. Contribute to the understanding of useful approaches for other NIH research training evaluation efforts.

1.3 Overview of the Feasibility Study Report

Following this Introduction, **Section 2** describes the Feasibility Study questions, methods, sources of data, and data collection approaches used. **Section 3** outlines the history, aims, and structure of the seven trans-NIH Summer Research Training Institutes.

The next three sections summarize the data and findings from the Feasibility Study. The discussions in these and subsequent sections interweave the findings and results in order to produce a broad logical framework that will ultimately support the outcome evaluation. **Section 4** addresses the structure and implementation of the seven trans-NIH Summer Research Training Institutes, while **Section 5** discusses the training institutes' outcomes and how the training institutes are believed to achieve them. It also focuses on the feasibility of comparison groups based on the perspectives of Federal stakeholders and faculty members who have taught at the training institutes.

Drawing upon these results, **Section 6** presents final recommendations, building on answers to the seven Feasibility Study evaluation questions.

II. Feasibility Study Questions, Methods, and Data Sources

This section describes the evaluation questions examined by the Feasibility Study and discusses the methods and data sources used to address these questions.

2.1 Feasibility Study Evaluation Questions

The primary goals of the Feasibility Study were to determine: (1) whether an outcome evaluation of the seven trans-NIH Summer **Research Training Institutes was** feasible and warranted; (2) the type of evaluation approach most appropriate for the outcome evaluation; and (3) the most appropriate evaluation design (or designs), key variables, data sources and data collection approaches given constraints on the client's time, budget, and available data. In order to address these goals, the evaluation team formulated seven specific evaluation questions as shown in Exhibit 1.

To answer these questions, the evaluation team followed a series of methodological steps that involved collection of data from

Exhibit 1. SRTI Outcome Evaluation Feasibility Study Questions

- 1. Is an outcome evaluation of the seven trans-NIH Summer Research Training Institutes Program feasible and warranted at this time?
- 2. What are the study questions an outcome evaluation should address?
 - 2a. What study questions should pertain to all of the training institutes?
 - 2b. What study questions may be unique to only some of the training institutes?
- 3. What are the appropriate outcome measures and metrics for evaluating the training institutes?
- 4. What is the optimal design (or designs) for evaluating these training institutes?
- 5. What types of comparison groups are feasible to address the study questions?
 - 5a. Is it feasible to use applicants and/or participants from one training institute as a comparison group for another training institute?
 - 5b. Is it feasible to use propensity score matching techniques to create a comparison group from unselected applicants?
- 6. What are the available data sources, and what are their strengths and limitations?
- 7. What new data collection instruments will be needed?

multiple sources. The methodological steps are outlined in Exhibit 2.

Exhibit 2. Methodology for the SRTI Outcome Evaluation Feasibility Study

- 1. Clarification of study objectives, issues, and questions
- 2. Focused review of the relevant literature
- 3. Interviews with relevant stakeholders (OBSSR staff members, NIH program officers, SRTI faculty members)
- 4. Identification of existing sources of data (e.g., training institute satisfaction surveys, recruitment files, participant lists, IMPAC II)
- 5. Development of a preliminary conceptual framework
- 6. Development of evaluation questions and key variables
- 7. Analysis of findings and recommendations for the outcome evaluation design

2.2 Data Sources and Data Collection Approaches

The data sources included analysis of documents, web pages, and data on past training institute trainees and unselected applicants; a literature review on current theoretical frameworks used in the evaluation of past training programs and previous evaluations of NIH-sponsored summer research training institutes; interviews with NIH program staff members involved in planning and developing the SRTIs; and interviews with a sample of faculty who had taught at the training institutes. In addition, the evaluation team conducted two pilot studies: one involving an analysis of trainee institute satisfaction questionnaires from the Institute on Systems Science and Health (ISSH), and a second study involving a comparison of 2011 trainees and a random sample of unselected applicants to the Training Institute on dissemination and Implementation Research in Health (TIDIRH). **Exhibit 3** shows how each of the data sources and collection approaches addressed specific Feasibility Study questions.

Feasibility Study Questions	Data Sources	Data Collection Approaches
 Is an outcome evaluation of the seven trans-NIH Summer Research Training Institutes Program feasible and warranted at this time? 	 Opinions of primary stakeholders Web pages and other materials provided by institutes Data on trainees and unselected applicants 	 Interviews with NIH stakeholders Interviews with SRTI faculty Data and document review
 What are the study questions an outcome evaluation should address? What study questions should pertain to all of the training institutes? 	 Opinions of primary stakeholders Web pages and other materials provided by institutes Review of evaluation studies on prior NHA sponsored training 	 Interviews with NIH stakeholders Interviews with SRTI faculty Document review
2b. What study questions may be unique to only some of the training institutes?	institutes	
3. What are the appropriate outcome measures and metrics for evaluating the training institutes?	 Opinions of primary stakeholders Review of evaluation studies on prior NIH-sponsored training institutes Review of training evaluation literature 	 Interviews with NIH stakeholders Interviews with SRTI faculty Document review
4. What is the optimal design (or designs) for evaluating these training institutes?	 Review of evaluation studies on prior NIH-sponsored training institutes Review of training evaluation literature Opinions of primary stakeholders 	 Document review Interviews with NIH stakeholders Interviews with SRTI faculty
5. What types of comparison groups are feasible to address the study questions?	 Review of evaluation studies on prior NIH-sponsored training institutes Pilot studies to assess feasibility of current data sources Opinions of primary stakeholders 	 Document review Analysis of grant and publications activity of trainees and unselected applicants Interviews with NIH stakeholders Interviews with SRTI faculty

Exhibit 3. Feasibility Study Questions, Data Sources and Data Collection Approaches

Feasibility Study Questions	Data Sources	Data Collection Approaches
5a. Is it feasible to use applicants and/or participants from one training institute as a comparison group for another training institute?	 Pilot studies to assess feasibility of current data sources Opinions of primary stakeholders 	 Analysis of grant and publications activity of trainees and unselected applicants Interviews with NIH stakeholders Interviews with SRTI faculty
5b. Is it feasible to use propensity score matching techniques to create a comparison group from unselected applicants?	 Review of propensity score matching literature Review of data on trainees and unselected applicants 	 Data and document review
6. What are the available data sources, and what are their strengths and limitations?	 Web pages and other materials provided by institutes Data on trainees and unselected applicants 	Data and document review
What new data collection instruments will be needed?	 Final results generated from data and documents 	Data and document analysis and synthesis

The following provides a brief description of the document review, literature review, federal and faculty interviews, and the two pilot studies.

2.2.1 Document Review

The evaluation team gathered background information on the structure of each SRTI through a review of existing documents and administrative records (shown in **Exhibit 4).** The document and record review generated considerable data. To organize and store the data in a useful format, the evaluation team created an Institute Data Template. The template provided a user-friendly organizational format that allowed the data to be described

Exhibit 4. Document and Administrative Records Included in the Document Review

- Web pages describing each of the SRTIs
- Published articles and internal reports provided by members of the Evaluation Advisory Committee (EAC)
- PowerPoint presentations by EAC members at the November 7, 2013 project meeting
- Spreadsheets containing data on trainees, unselected applicants, and faculty members for the SRTIs

and identified by document source for each SRTI. The template contained the following categories of information: institute overview (history, goals, etc.); faculty; trainees (target population, selection criteria, application materials and selection process); institute structure and curriculum; any prior evaluations of the training institute; and the type and availability of data on trainees and unselected applicants indexed by year.

2.2.2 Literature Review

The evaluation team conducted a comprehensive review of relevant literature on previous evaluations of NIH training institutes, other related training programs, and related studies. The review included both published and unpublished literature as well as a search of the NIH Evaluation Set-Aside Program's unpublished reports of

evaluations of NIH programs. In conducting the review, the evaluation team sought to answer three broad questions (shown in **Exhibit 5**).

The literature search was conducted in November and December 2013, using Scopus, PubMed, Google, and Google Scholar. Search terms included: summer research training institute, summer training institute, research training, training evaluation, training impact, and Kirkpatrick Four Levels® Model. Searches were supplemented through manual review of references from relevant articles, and of tables of contents from selected journals. This so-called "snowball approach" was used with evaluation journals such as the American Journal of Evaluation. Evaluation and Program Planning,

Exhibit 5. Literature Review Questions

- 1. What conceptual frameworks and models have been used in the evaluation literature to evaluate training programs, and what are their strengths and weaknesses?
- 2. How have federally-funded summer research training institutes been evaluated in the past? What can these studies tell about:
 - (a) The types of process and outcome variables used in such studies?
 - (b) The extent to which comparison groups have been used and how such groups were selected and constituted?
 - (c) The existence of any widely accepted and utilized measures or instruments?
- 3. What implications do the findings from this review have for the design of the SRTI outcome evaluation?

New Directions in Evaluation, and Evaluation in the Health Professions. All searches looked for review articles and for specific evaluation reports on federally-funded summer research training institutes. These searches yielded 103 articles including reviews and evaluations of various training institutes. The evaluation team defined a summer research training institute as a NIH-sponsored short-term residential learning experience (less than 30 days) that focused on biomedical research topics and were open to individuals at the postgraduate level or higher. This is a relatively narrow definition designed to rule out evaluations of programs at universities, or training that focused on issues other than research knowledge and skills.

Applying these search criteria, the evaluation team identified a total of 18 published reports describing results from 21 specific evaluations of federally-sponsored summer research training institutes that met our criteria. We also identified a total of 18 published and unpublished review articles, discussions, training handbooks, and other relevant reports that discussed theoretical models applicable to evaluation of research training programs.

2.2.3 Interviews with Federal Stakeholders and Training Institute Faculty

Stakeholder interviews were critical to fully understanding the initial purpose, development, and inner-workings of the SRTIs. The evaluation team identified two key stakeholder groups. The first group consisted of NIH program staff members from various NIH Institutes and Offices that had participated in planning and developing the SRTIs. A total of 17 NIH program staff members were identified and 16 were interviewed (94% response rate). Program staff members were employed by seven NIH Institutes and Offices, including OBSSR, the National Cancer Institute (NCI), the National Heart, Lung, and Blood Institute (NHLBI), the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD), the Office of Disease Prevention (ODP), the National Institute of Mental Health (NIMH), and the National Institute on Alcohol Abuse and Alcoholism (NIAAA). Two stakeholders no longer worked with the federal government at the time of the interviews. All seven SRTIs were represented by at least two federal stakeholder interviews.

The evaluation team developed an interview protocol which included questions about the history of SRTI, its structure and implementation (including the curriculum, teaching methods, the role of mentors) and perspectives on the types of evaluation outcomes and questions each respondent believed to be most important for the outcome evaluation. One three-hour interview involved two respondents who were interviewed in person. The evaluation team conducted the remaining interviews by telephone. These interviews ran from 47 to 100 minutes in length, and averaged 69 minutes (excluding the three-hour interview). The interviews were conducted during April 2014 using two-person teams, with one individual serving as the lead interviewer and the second as a note-taker. Interviewers were recorded as an additional resource for the note-taker, and copies of these recordings were destroyed at the conclusion of the project.

The second group of stakeholders consisted of SRTI faculty members with university or other non-federal organization affiliations. These faculty members taught at one of the SRTIs as either a guest faculty member or as a core faculty member who helped to develop the training institute. A total of seven faculty members were interviewed. Although a representative from each SRTI was attempted, only six SRTIs had faculty that were available in the targeted time frame. The Institute on Systems Science and Health (ISSH) consisted of three different tracks and this institute was represented by two interviews. All seven interviews occurred in June 2014 and covered similar content to the federal stakeholder interviews with additional issues specific to faculty only, including the faculty member's role as either a core or guest faculty member, the curriculum they developed and taught, their assessment of participant's response to the curriculum, and their experience as a host institution if applicable. Faculty Interviews required between 42 and 86 minutes with the interviews lasting 57 minutes on average and were conducted by telephone only. Responses were recorded and tabulated using the same methodology as the federal stakeholder interviews.

2.2.4 Pilot Studies

The evaluation team conducted two pilot studies to investigate the feasibility of conducting effective evaluation research using existing SRTI data. The first pilot study analyzed participant satisfaction survey data collected from the three ISSH learning tracks over a four year period. The second pilot study analyzed grant and publication activities from participating and non-participating applicants of the TIDIRH institute.

ISSH Pilot Study

The purpose of the ISSH pilot study was to analyze existing post-training participant satisfaction data to determine the level of depth and detail to assess usefulness in measuring institute outcomes. A total of 200 participant satisfaction forms were collected by the track leaders of four ISSH institutes held in 2009, 2010, 2011, and 2012. The forms specifically collected participant satisfaction data from each of the

three specialized learning tracks offered at this institute: 1) 62 forms were collected from the Network Analysis Track; 58 forms were collected from the System Dynamics Track; and 80 forms were collected from the Agent-Based Modeling (ABM) Track. The analysis focused on the open-ended survey items, which covered such areas as:

- Feedback about what the track did well
- Feedback about what the track could do better
- Level and type of collaborations formed
- Satisfaction with the length of the track and use of track time

These items were analyzed to determine the depth and detail of the responses in order to gauge the usefulness of the forms in gathering detailed feedback about the quality of the ISSH institute tracks. Each track was analyzed separately.

TIDIRH Pilot Study

The TIDIRH pilot study analyzed applicant and participant data from the 2011 TIDIRH institute. The purpose of this pilot study was to determine whether or not data existed in large enough numbers with enough detail to allow for comparison group testing. The dataset from the 2011 TIDIRH institute was selected because of its size (over 250 applicants), the availability of non-participant data, and the amount of time that had passed since the institute (2011) was adequate for checking grant and publication progress in early 2014. To conduct the study, 66 individuals were chosen from the applicant list. Half (33) were participants in the 2011 institute. The other 33 were randomly selected from the list of non-participants who had applied to the institute (every seventh applicant was selected from the total list of 264 applicants, not counting participants). Once the sample of 66 was selected, searches were conducted using the *IMPAC II QVR* person reports database. The searches were conducted in January 2014. The database was searched for:

- The number of grants submitted and awarded since the 2011 institute, beginning on or after January 2012
- The total number of publications
- The total number of first author publications
- The number of publications since the 2011 institute
- The number of first author publications since the 2011 institute
 - Selected publications included items listed with publication dates on or after January 2012
 - PubMed searches were completed for individuals not listed in QVR

Data were downloaded so that basic statistical comparisons could be conducted (chi square, t-test).

III. The Trans-NIH Summer Research Training Institutes

This section begins the discussion of the results of Feasibility Study data collection and includes a brief overview of the history of the SRTI program as a whole, and the history, aims, and structure of each SRTI, emphasizing the specific problem (or problems) each was created to address. The information for the overviews of each training institute was taken primarily from the document review with some additional background from the interviews with the NIH Program Officers.

3.1 History of the Trans-NIH Summer Research Training Institutes Program

A fundamental goal of OBSSR is to foster and promote a behavioral and social sciences interdisciplinary perspective within NIH's research portfolio. This goal is especially significant because approximately 40% of premature deaths in the United States are related to mutable behavioral factors such as smoking, poor diet, stress, inactivity, violence, and accidents, as well as broader societal factors including poverty, discrimination, gender biases, lack of medical insurance, and poor access to quality health care. Furthermore, behaviors such as overeating and reduced physical activity have resulted in an obesity epidemic in today's youth that threatens to cut average lifespan for the first time. Behavioral and social factors also have important implications for other diseases. For example, in people with chronic pain conditions, behavioral interventions are as effective as pain killers (and non-addictive), while specially designed diet and exercise programs are more effective in reducing the onset of diabetes in people at-risk than existing medical treatments. These findings and others highlight the need for research attention to these behavioral and social factors to improve health, quality of life, and increased longevity for all.

Research on these behavioral and social factors in the biomedical domain requires scientists with specific skills and knowledge grounded in both the behavioral and social sciences and the biomedical field. The complexity of the problems posed in biomedical research requires a multidisciplinary research team that can combine skills and knowledge across these fields. Biomedical and behavioral and social scientists lacked familiarity with each other's fields and were not trained in working together as a team. At the time OBSSR was established within NIH in 1999, it was quickly evident that research proposals bridging these fields were not faring well in the internal grant review process. There was a clear need to provide learning and training experiences for both groups of research investigators that would familiarize each group with the others' capabilities and knowledge, increase skills and knowledge in specific topical areas, and teach the specialized skills needed to work effectively in multidisciplinary research teams.

Beginning in FY 2001, OBSSR sponsored its first summer research training institute on *Randomized Clinical Trials Involving Behavioral Interventions (RCT)*. This training institute provided a brief (nine-day) residential educational experience for a small number of investigators with either behavioral and social sciences or biomedical research training. From its inception, the RCT was a highly popular training institute.

OBSSR subsequently partnered with program staff from the National Heart, Lung, and Blood Institute (NHLBI) to offer this institute on an annual basis.

The growing popularity of the RCT institute and anecdotal indications of its success in training researchers to design and conduct higher quality randomized clinical trials involving behavioral interventions led OBSSR to launch additional training institutes in partnership with program staff from other NIH Institutes, and some outside federal agencies and private organizations. The first four training institutes addressed well-established areas of scientific inquiry, including the application of health behavioral theories, social and behavioral interventions, and applied child and adolescent development research. Beginning in FY 2009, OBSSR and its partners funded training institutes on rapidly emerging scientific fields that showed high promise for application to biomedical problems. These fields included dissemination and implementation research, the use of mobile health technologies in health care, and the application of systems sciences in health. The goals of these latter three institutes differed somewhat from the institutes offered on established scientific areas; an important additional goal of the former group was to develop a sustainable research community in these fields while developing a productive research portfolio.

Between FY 2001 and FY 2013, OBSSR has provided a total of \$15 million dollars in funding to support the seven training institutes shown in **Exhibit 6**.

Program	Established or Emerging Fields	First Year Offered	Estimated Annual Cost	Cumulative Total
Randomized Clinical Trials Involving Behavioral Intervention (RCT)	Established	2001	\$470K	\$5.6M
NIH Summer Institute on Social and Behavioral Intervention Research (SW)	Established	2004	\$350K	\$2.8M
Advanced Training Institute on Health Behavior Theory (ATI)	Established	2004	\$300K	\$2.7M
Summer Institute in Applied Research in Child and Adolescent Development (C&AD)	Established	2007	\$250K	\$750K ³
Institute on Systems Science and Health (ISSH)	Emerging	2009	\$425K	\$1.7M
Summer Training Institute on Dissemination & Implementation Research in Health (TIDIRH)	Emerging	2011	\$350K	\$700K
mHealth Training Institute (mHealth)	Emerging	2011	\$350K	\$700K
Total			\$2.5M	\$15M

Exhibit 6. The Seven Summer Research Training Institutes

In recent years, funding has become more restricted and there have been discussions about whether the need for these training institutes continues to exist and whether OBSSR should continue to fund them (versus encouraging universities to

³ Summer Institute in Applied Research in Child and Adolescent Development is held every other year.

develop their own training initiatives in these areas or offering the content through online training). These issues have led to an interest in evaluating the seven training institutes in order to better understand what has been accomplished through them. That interest resulted in the commissioning of this evaluation Feasibility Study.

3.2 History of the Seven Summer Research Training Institutes

This section presents a brief historical overview of each of the seven training institutes. Throughout this Feasibility Study, the evaluation team has engaged in a discussion of whether the seven training institutes collectively represent a program at OBSSR or whether each institute arose separately in response to specific needs and their championing by OBSSR and other NIH staff. We found support for both points of view, and agreed that, taken collectively, the seven institutes share common core components and at least some programmatic goals. The overview of the seven training institutes is presented in order of each institute's creation.

3.2.1 Randomized Clinical Trials Involving Behavioral Interventions (RCT)

As noted above, the RCT was created in FY 2001 to address a specific problem. Within the biomedical research field, randomized clinical trials involving behavioral interventions were controversial, in part because many of the outcome measures such trials require were perceived to lack the precision (and even validity) of outcomes used in traditional biomedical research. At the time the RCT was developed, there were several examples of trials involving behavioral interventions that had been poorly designed, and grant applications proposing such studies were not scored favorably in the internal grant review process. OBSSR program staff were aware of these problems and discussed the idea of developing a residential learning experience that would provide junior research investigators (viewed as those in the first five years of their research careers) with training on the mechanics of planning, designing, and conducting randomized clinical trials with behavioral interventions. Junior investigators were identified as the target population for this training because they were perceived as more open to learning new knowledge and skills and because training investigators at an early stage in their careers would ensure that they could build research careers by pursuing grants in this area. RCT core curriculum topics are listed in Exhibit 7.

The RCT was the first training institute contract issued by OBSSR, and has been taught annually for the past 13 years. It was based on an earlier institute one of the two OBSSR co-founders had attended that was called the *Epidemiology and Prevention of Cardiovascular Disease*. This training institute was offered on an annual basis since the 1970s at a resort at Lake Tahoe.

The RCT is a nine-day residential training program held annually at the Airlie Conference Center near Warrenton, Virginia. The location was chosen because it is an isolated setting

Exhibit 7. RCT Core Curriculum Topics

- Principles underlying conduct of unbiased clinical trials
- Unique challenges of behavioral interventions
- Evaluate RCT designs in terms of their appropriateness for scientific and clinical goals
- Selection of appropriate strategies for enrollment, randomization, and retention
- Monitoring, coordinating and conducting clinical trials
- Developing analytic strategies
- Evaluate reports of behavioral RCTs
- Design an RCT with work group

free from distraction; this isolation encourages immersion and internal focus by the community of trainees and faculty who attend. To date, about 468 trainees have completed the RCT training institute.

3.2.2 NIH Summer Institute on Social and Behavioral Intervention Research (SW)

The NIH Summer Institute on Social and Behavioral Intervention Research (SW) training institute was part of a broader NIH response to a Congressional mandate to increase the involvement of social work research investigators in biomedical research at NIH. In 2003, a trans-NIH planning group issued a report with several recommendations for enhancing social work research investigators' participation in biomedical research. This report (the NIH Plan for Social Work Research) recommended creating a specific summer research training institute for social work researchers based on the recognition that many newly-minted social work PhD investigators moved directly from doctoral training in graduate school to assistant professor appointments in schools of social work without the benefit of post-doctoral fellowship training and mentoring in research methods. The summer research training institute was intended to partially fill this gap by offering training on the conceptual, methodological, and practical issues involved in planning and conducting research on the impact of behavioral and social interventions on health outcomes, health behavior, and treatment. As envisioned by the 2003 plan, the training institute would include lectures, seminars, and small group discussions on research design and methods, discussion sessions on methodological approaches and interventions, and consultation on the development of research interests and advice on preparing and submitting research grant applications to NIH. OBSSR partnered with NCI, NIAAA, NIMH, NICHD, the National Institute on Aging (NIA), and the National Institute on Drug Abuse (NIDA).

An important difference between this SRTI and the other six is that while the main emphasis on training in research design and methods remained constant, the context in which these would be applied changed. Thus, the first year of the SW training institute addressed qualitative and mixed methods research, the second year targeted behavioral and social science interventions, and later years included design and development of quantitative research in social work (two years), cross-systems research to improve health outcomes, community based participatory research, transdisciplinary research (genetics and social work), and behavioral and social interventions.

The NIH Plan for Social Work Research also promoted the development of specific Funding Opportunity Announcements (FOAs) that were designed to encourage research grant applications on social work research in health. FOAs were issued in FY 2006 and FY 2007 for this purpose.

3.2.3 The Advanced Training Institute on Health Behavior Theory (ATI)

Planning for the ATI summer research training institute began in FY 2002 as program staff at NCI and leading cancer behavioral sciences research investigators recognized that there was a strong need to strengthen the quality of grant applications (and research more broadly) involving the use of health behavior theories. NCI partnered with the (then) National Center for Complementary and Alternative Medicine (NCCAM), the National Institute of Allergy and Infectious Disease (NIAID), the National Institute of Dental and Craniofacial Research (NIDCR), NIDA, NHLBI, and OBSSR to offer the first ATI institute in 2004. Six training institutes have been held to date (2004, 2005, 2006, 2008, 2010, and 2012) and about 150 trainees have completed the eight day session.

The ATI training curriculum is organized around four basic themes: fundamental issues in health behavior theory; current theories and types of theories; theory testing; and application of health behavior theories to health behavior change interventions. The ATI is a seven-day institute that targets new or early-stage investigators. The training institute seeks to train participants to think differently about their world and what it means to use theory. ATI targets new and early stage investigators (especially post-doctoral fellows) and tends to attract learners from two groups: those who were trained in behavioral and social science research and want to improve their capacity to apply these theories in the biomedical field; and individuals with biomedical training who want to learn about how to use behavioral and social science theories in their work.

The ATI institute was initially offered at La Jolla, California, but in subsequent years has been hosted at the Fluno Center at the University of Wisconsin-Madison.

3.2.4 Summer Institute in Applied Research on Child and Adolescent Development (C&AD)

The C&AD is sponsored by NICHD in collaboration with OBSSR and the Society for Research in Child Development. Established in FY 2006, the C&AD institute seeks to help new and early stage research investigators develop advanced skills in using theory and appropriate methodology for developmental practices, interventions, and/or materials that improve the lives of children, support investigators who are beginning careers as applied researchers by providing training that will build on their existing content knowledge and research skills, and provide mentoring and guidance on topics including alternative funding support, dissemination of findings, and policy-relevant research. An important part of the training institute is work that each trainee does on an initial research grant application.

When trainees are selected, they submit specific questions they have concerning applied research in child and adolescent development. These questions are then used as a framework by the faculty for each year's training institute. In this way, the C&AD provides a basic core curriculum but fine-tunes it to reflect the specific questions and problems faced by that year's trainees.

The C&AD was first offered in FY 2007. Participating trainees from that training found the various lectures highly valuable, and NICHD made the decision to publish these lectures as a sourcebook⁴ for future trainees. In addition to the sourcebook, there is a Facebook page where past and current participants can interact before, during, and after the training institute sessions.

⁴ Maholmes, V. and Lomonaco, C.G. (Eds.) *Applied Research in Child and Adolescent Development: A practical Guide*. New York: Taylor and Francis, 2007.

3.2.5 Institute on Systems Science and Health (ISSH)

ISSH was the first of the three later SRTIs to address an emerging scientific field of inquiry. In 2005 and 2006, senior staff within OBSSR grew interested in exploring whether and how research methods that were then finding acceptance in engineering, information science, and computer science could be adapted to the study of health and biomedical problems. This interest led to a general and informal review of various methods and an assessment of which ones were ready for expansion into the health and public health fields. One individual at OBSSR championed the promotion of these methods and teamed with a second individual at the Centers for Disease Control and Prevention (CDC) who was obtaining his doctorate in systems science. Three systems science methodologies were identified as appropriate fields with which to begin: social network analysis (which was already finding strong acceptance in sociology); agent-based modeling; and systems dynamics modeling.

In FY 2007, OBSSR sponsored a Symposium on Systems Science in Health. This symposium brought together several prominent investigators from these three scientific fields who shared their ideas on how these methodologies could be applied in research on health and biomedical problems. The symposium was video-recorded and the recording was made available as a training resource by NIH within the United States and internationally, where it was highly viewed. The strong and positive reception of investigators in the United States and abroad led OBSSR and CDC to partner on the development of a summer research training institute intended to acquaint behavioral and social science and health researchers with the three systems methodologies and educate investigators from engineering and other fields who were familiar with these methods on the kinds of health and biomedical problems for which their application would be appropriate. The first Institute on Systems Science and Health was held in 2009.

The week-long training is designed to provide trainees with a thorough introduction to one of the three systems science methodologies and to show how they can be used to study complex dimensions of medical and public health problems. ISSH is the only SRTI organized into separate tracks (one for each of the three methods). Trainees apply for and participate in a single track focusing on that specific method. The five-day training institute also contains plenary sessions that all trainees attend; these sessions ensure that trainees are aware of the other two methods while concentrating their time and energy on one. Trainees learn the basic principles of the systems method, how to use software supporting that method, and work on individual research projects utilizing that method. An important part of the institute is the opportunity to meet with NIH program staff from different institutes who provide information on research opportunities and NIH priorities.

The ISSH is part of a broader effort to build a research community (and research portfolio) in the systems science field. In addition to the training institute, there was a listserv that was developed beginning in 2007 to provide former trainees and others who are interested in the systems sciences with an electronic commons where they can meet and exchange ideas, contacts, resources, etc. A third major component of this

effort is the issuance of several Funding Opportunity Announcements that specifically encourage research applications applying systems science methods in health research.

3.2.6 mHealth Training Institute (mHealth)

Another emerging scientific area is the rapidly growing field of mobile and wireless health technology, which offers exciting potentials for the real-time collection of continuous biological, behavioral, and environmental data. These technologies provide a means of continuously monitoring chronic diseases in people from all over the world. The integration of these technologies into health care research and tools has been slower to occur, and designing and conducting rigorous evaluations of their effects will require the development and application of alternative research design methods that can be matched to the appropriate stage of each technology's evolution.

The mHealth training institute was first offered in June 2011 with a goal of building research capacity by familiarizing health researchers with the technologies and the issues they stimulate. At the same time, OBSSR recognized that the technology community also needed greater familiarity with the issues, requirements, and tools of biomedical and health research. During the first year, OBSSR partnered with QualComm, a private technology firm, but several other outside federal agencies (the National Science Foundation (NSF) and the Food and Drug Administration (FDA) had input on the planning as well as providing some faculty. The five-day training institute brought together leaders in mobile technology and the behavior sciences in order to provide this cross-training for both groups (Nilsen et al., 2012). The curriculum covered the current state of mobile and wireless technology and engineering, behavior change theories and their application, and new developments in evaluation design and data collection and analysis. The training institute lectures were video-recorded and made available through the OBSSR mHealth website. Subsequent mHealth training institutes were held in December 2011 and in the summers of 2012 and 2013.

As was the case for systems sciences, OBSSR sought to create an informed research community that could strengthen the NIH research portfolio in this emerging field. In August 2011, OBSSR co-sponsored a mHealth Evidence Workshop to identify the best strategies for generating evidence of the efficacy and effectiveness of mobile health technologies. In addition, OBSSR established a listserv to provide information on upcoming events, research opportunities, and other news. Unlike ISSH, however, no specific NIH FOAs were created to support research on mHealth. (The Agency for Healthcare Research and Quality recently issued a Request for Applications on mobile health technologies.).

3.2.7 Training Institute on Dissemination and Implementation Research in Health (TIDIRH)

The TIDIRH web page notes that "one of the most critical issues impeding improvements in public health today is the enormous gap between what we know can optimize health and healthcare and what actually gets implemented in daily practice. The science of dissemination and implementation (D&I) seeks to address this gap by understanding how to best ensure that evidence-based strategies to improve health and prevent disease are effectively delivered in clinical and public health practice." NIH has been actively involved in promoting D&I research and developing research capacity in this field since the early 2000s (Glasgow et al., 2012). Several NIH Institutes have issued FOAs and generated research portfolios on D&I research issues. OBSSR convened a trans-NIH committee to promote D&I research in 2003. Beginning in 2005, NIH issued several trans-NIH program announcements for D&I research. Initially co-sponsored by eight Institutes, OBSSR, and the Office of Dietary Supplements, these FOAs resulted in the funding of 40 D&I grants. When the program announcements were reissued in 2009, four additional Institutes and the Fogarty International Center joined the original group.

As one step in the overall process of building research capacity in the D&I field, OBSSR, NCI, NIMH, and the Department of Veterans Affairs (VA) co-sponsored the first Training Institute in Dissemination and Implementation Research in Health in August 2011 (Meissner et al., 2011). The five-day institute has two broad goals: to prepare research investigators to conduct D&I research, and to stimulate interest in D&I research at institutions around the United States. The training institute targets both new and experienced investigators who have not previously been funded to conduct D&I research. Key objectives of TIDIRH include increasing the submission rate and quality of D&I grant applications and publications, and the use of a "train the trainer" model in which trainees are expected to return to their home academic institutions and teach what they have learned to their students, post-doctoral fellows, and fellow faculty members. As a means of building interest in D&I research among research institutions, the site of the training is rotated among interested potential host institutions each year.

The curriculum covers a broad range of topics, although a major focus of the fiveday training is each trainee's work on refining and improving an individual research proposal. Morning lectures address topics such as theory, implementation and evaluation approaches in D&I research, creating partnerships and multi-level transdisciplinary research teams, research design, methods, and analytical techniques appropriate to D&I research, and conducting research at different and multiple levels of interventions. After the first year, the curriculum included a stronger emphasis on creating successful transdisciplinary research teams. Afternoons provide time for small group work and refinement of the individual research proposals each trainee brings to the institute.

3.3 Key Findings for the Seven Summer Research Training Institutes Based on Document Review and NIH Program Officer Interviews

The overview of the seven SRTIs shows that they share several important similarities and differences, including the following:

 The SRTIs address different scientific fields and methodological issues. Four of these fields are well-established scientifically (e.g., applied research in child and adolescent development, health behavioral theories, randomized controlled trials with behavioral interventions, and social work health-related practice). Three institutes address recently emerging fields of inquiry (dissemination and implementation research, mobile health technologies, and systems science). While a common goal across all seven training institutes was to increase the grant submission rate and quality for research grant applications and publications), the three institutes addressing emerging fields also sought to build research capacity and a community of active researchers.

- The three emerging training institutes involved other federal partners, including the CDC, the VA, FDA, and NSF. Only one of the four training institutes that addressed established fields had a non-NIH partner—the C&AD training institute, for which NICHD partnered with OBSSR and the Society for Research in Child Development.
- The seven training institutes have been offered for varying lengths of time. The four institutes addressing established fields were the first ones offered, beginning as early as 2001 (RCT). The three institutes addressing emerging fields were begun more recently (2009 and 2011).
- The specific problems that led to the creation of the SRTIs differed and these differences affected the types of trainees targeted by each institute. For the institutes addressing established fields or methods (RCT, ATI, and C&AD), the perceived problem was that research investigators had not received sufficient training in these fields during their graduate education. In the case of the SW training institute, the problem was viewed as structural—schools of social work were hiring junior faculty directly after graduation, and these investigators were not entering into postdoctoral fellowships where they might have received additional training and mentoring to equip them for research careers. All four of these institutes targeted new or early-stage investigators who would presumably advance after training to full research careers focusing on these fields. For the three institutes addressing emerging fields, the fields themselves were so new that no one had received much training in them. Because one of the goals of these institutes was to build research capacity and create an active research community in each field, each institute sought to recruit both new and experienced investigators who could work quickly to apply for new grants and build research portfolios in the targeted fields.
- A critical difference between the institutes addressing emerging versus established fields is that the emerging initiatives were embedded in broader initiatives within NIH aimed at promoting these scientific fields. These broader initiatives included components such as specific research funding opportunities, listservs that could promote connection and collaboration among former participants, and trans-NIH planning groups or committees that raised awareness of these emerging fields. In addition, the emerging fields training institutes all had non-NIH federal and/or private partners who co-sponsored the training with OBSSR.

Based on the data from this review of documents and records, we constructed a broad conceptual model relating the seven Summer Research Training Institutes to OBSSR's overall mission (**Exhibit 8**).



Exhibit 8. Conceptual Model for the Trans-NIH Summer Research Training Institutes Program

IV. Structure and Implementation of the Summer Research Training Institutes

This section describes the structure and implementation of the seven trans-NIH Summer Research Training Institutes. The overall implementation process can be divided into two smaller sets of activities: (1) planning and development of the SRTIs, and (2) the recruitment and selection of faculty, the host institution, and trainees. The description of these activities draws upon data from the NIH Program Officer and training institute faculty interviews. This section concludes with an Implementation Model that visually summarizes the implementation theory of the seven Summer Research Training Institutes. The implementation theory describes *how the training institutes are intended to work*.

4.1 Planning and Development of the Summer Research Training Institutes

The results from interviews with the NIH Program Officers and training institute faculty indicate that planning and design of each of the seven training institutes drew heavily upon basic principles of adult learning. While different authors have stated these principles in different ways, these statements generally appear similar to those stated by Malcolm Knowles (1996):

- Adults are internally motivated and self-directed.
- Adults bring their life experiences and prior knowledge to learning experiences.
- Adults are goal-oriented.
- Adults are relevancy-oriented.
- Adults are practical.
- Adult learners like to be respected.
- Adults learn best by experiencing a blend of activities that promote the three learning domains (cognitive, affective, and behavioral).

These principles mean that trainees bring to training experiences their own personal learning objectives. While these vary widely, interviewees told us that most trainees wanted to participate in the training institutes because they were highly interested in the specific fields and methods addressed by the institutes and because they wanted to increase their success in their research careers. The basic adult learning principles also indicate that trainees would value training that was provided using a variety of instructional approaches (e.g., lecture, small group, informal discussions, individual projects, etc.). Both groups of interviewees described a continual need in the first years of each training institute to strike a balance between didactic and interactive methods. This generally involved reducing time spent in lectures. For the multi-track ISSH institute, some fine-tuning was needed to enable trainees enrolled in one track to learn something about the methods taught in the other two tracks.

4.1.1 Designing a Curriculum

The initial curriculum developed for each training institute's first year was designed by an advisory committee. Usually the advisory committee was composed of NIH Program Officers from the partnering NIH Institutes, although some committees included outside representatives and in one case (C&AD) the committee consisted exclusively of outside investigators who were leaders in their fields. Several training institutes involved outside federal partners (e.g., those addressing emerging fields). After the initial year, each training institute incorporated feedback from trainees (and faculty) on what lectures were especially successful and which ones were not. In some cases, this feedback also identified additional topics that were included in the following year's curriculum.

While the specific content of each training institute's curriculum was unique to that institute, there were several common topics covered across all seven training institutes as shown in **Exhibit 9**). The training curricula always included material on theory and methods in the topical areas, and several training institutes also included material on grantsmanship, working on multidisciplinary research teams, and building a new research field.

Training Institute	Theory and Methods	Grantsmanship	Working in Multidisciplinary Teams	Building a New Research Field
ATI	✓			
C&AD	✓	✓		
ISSH	\checkmark	\checkmark	\checkmark	\checkmark
mHealth	\checkmark		\checkmark	\checkmark
RCT	\checkmark		\checkmark	
SW	\checkmark	\checkmark		
TIDIRH	\checkmark	\checkmark		\checkmark

Exhibit 9. Curriculum Content

Changes did occur to the training curricula over time as shown in **Exhibit 10**). For some training institutes, the general theme of the training differed from one year to the next (C&AD, SW).

Training Institute	Description of Changes to Curriculum
ATI	 Curriculum stayed relatively the same until 2010. In 2010, collaboration with NIDCR brought in a new topic area (dentistry). Curriculum strengthened emphasis on theory-based interventions and behavioral outcomes. Curriculum expanded to include emphasis on mediating and moderating variables.
CAD	 Themes for next year are based in part on what the developers learn from the participating trainees in the current year.
ISSH	• Basic systems theories and methods remain unchanged but each year's Track Leaders introduce new themes and teaching approaches.
mHealth	The curriculum has remained relatively the same over time.
RCT	• Core elements of the curriculum (about 2/3) stay relatively the same each year, but about 1/3 of content changes to reflect new developments and issues in the field (ethics,

Exhibit 10. Changes to Curriculum

Training Institute	Description of Changes to Curriculum
	 new statistical models, controversies in the field). After 6 years (~2005) there was an increased emphasis on development of behavioral interventions and the complexities of the intervention and control group.
SW	 Core elements of the curriculum remain relatively the same, but there is a different overall theme each year. There has been an increasing focus on the nature of the behavioral interventions.
TIDIRH	The curriculum has remained relatively the same over time.

For the other training institutes, the core content generally remained similar but information on new developments or controversies in a scientific field was included to update the curriculum. One trend that is evident among the training institutes that address established fields is the growing importance of the design and development of behavioral interventions.

Training institute faculty interview revealed that the faculty felt comfortable placing their own emphases on the topics they were asked to teach. For the most part, the training institutes accepted this and did not try to closely monitor what each faculty member would teach. One training institute did require a pre-approval of each faculty member's teaching materials in advance of the training institute.

4.1.2 Mentoring

All interviewees agreed that mentoring was a critical element for each of the seven training institutes. They noted that mentoring could be either formal or informal. Formal mentoring occurs when a trainee is assigned to a specific training institute faculty member with whom he or she works throughout the duration of the institute. When faculty members are assigned as mentors, the training institute developers attempt to match the trainee and the mentor on field of research interest. Informal mentoring occurs when trainees are not assigned a formal mentor but connect as a result of common research interests.

Five of the training institutes assign mentors to trainees either individually or by assigning the mentor to the small group within which the trainee will work during the training. Only two of the training institutes (ISSH and RCT) do not assign mentors; in these cases, several faculty members rotate among the different trainee groups. The assignment is not rigid, and considerable informal mentoring takes place as trainees and faculty with shared interests find each other during the training.

Mentoring can also occur between a trainee and his or her peers. The C&AD training institute assigns peer mentors based on shared research disciplines or interests prior to the date of the training institute so that trainees and peer mentors can contact each other and begin to become acquainted before the institute begins. C&AD encourages these peer mentoring relationships to continue after the training has ended and its planners have established a Facebook page and a Google Hangout to assist trainees with maintaining contact. How often this happens in reality is not known.

Three of the seven faculty interviewed indicated they have continued at least some of their mentoring relationships after the conclusion of an institute. One respondent discussed knowing that other faculty-trainee mentoring relationships had continued. On the whole, faculty stated there were no formal mechanisms for mentoring relationships to continue and it really depended on the personal relationship that had formed between the mentor/mentee at the training institute. They also suggested that the relationships tended to be driven more by the trainees reaching out to mentors after the training rather than from the mentor choosing to check in to see if there were any trainee needs. If trainees did not reach out, continued relationships were less likely to happen.

Several of the training institutes have established mechanisms for supporting ongoing contact between trainees and faculty after the training ended as shown in **Exhibit 11**. Maintenance of these mechanisms depends heavily on the ability of NIH program staff to devote the time and effort to keep them going.

Training Institute	Monthly Webinars	Listservs	Alumni Groups & Other
ATI			
C&AD		Google Hangout & Facebook pages	
ISSH		Listserv	Study groups
			Book clubs
mHealth	Monthly webinars	Listserv	
RCT		Failed Listserv	
SW			Get-togethers at annual meetings of professional organizations
TIDIRH	Monthly webinars	Listserv	

Exhibit 11. Post-Training Support Mechanisms

4.1.3 Individual Projects

One of the ways the training institutes incorporate the principles of adult learning into their designs is through the use of a specific individual or group project that provides a continuous mechanism for applying what trainees are learning over the course of the training institute. Trainees are required to prepare a research problem, idea, or concept paper as part of the application process. In most cases the quality of this individual project is reviewed as part of the trainee selection process.

The use of individual (or small group) projects as a teaching tool by the SRTIs is described below in **Exhibit 12**. The individual project often serves as a preliminary research grant proposal, and at some of the training institutes, the proposals are actually reviewed and feedback is provided to the trainees. For example, the SW training institute conducts a "mock NIH review" of some proposals as a means of highlighting the kinds of issues that reviewers are likely to raise.

Training Institute	How Individual Projects are Used in the Training Process			
ATI	 Trainees have one-on-one meetings with a faculty member to discuss their research problem/idea at the start of the week and again at the conclusion of the week. Trainees work together in small groups where they can present their research problem/idea and obtain feedback. At the end of the training institute, trainees complete an exercise in which they apply what they have learned to their individual research problem/idea. 			
CAD	 Trainees work on their research problem/idea throughout the institute; at the conclusion they present these ideas to federal program officers and receive feedback on them. 			
ISSH	 Selection for the training institute is based in part on how well-crafted the idea is and the rationale for applying the specific systems methodology they are trained in to the problem/idea. Over the course of the institute they work on developing their problem/idea using their assigned track. 			
mHealth	 The training institute uses Scenario-based learning, a team-based learning approach in which the multidisciplinary team works on a specific exercise over the course of the institute. Based on review of participant reaction forms, the institute developers learned that trainees were learning more from the group work than from the lectures, so they increased the former and decreased the latter. 			
RCT	• A recently instituted small-group exercise involves trainees reading a recently published report on a clinical trial, deciding what the next logical research question(s) should be, and designing a randomized clinical trial to address the research question(s).			
SW	 Trainees complete a series of assignments over the course of the institute in which they work on improving and refining various parts of their original concept paper and turn this into a grant application. At the end of the week there is a mock NIH review of some of the applications with feedback to the trainee. 			
TIDIRH	 Trainees work on refining and improving their concept papers over the course of the institute. The original model called for presentation of the papers at end of institute for feedback; however, it was difficult to go through 25-30 concept papers in one day. TIDIRH introduced the use of the Ignite PowerPoint presentation format in which each trainee has 3 minutes to present an idea using no more than 12 slides; slides change automatically every 15 seconds, therefore the presentation has to be very focused. 			

Exhibit 12. Individual Projects

Training institute faculty indicated that participants' projects were one of the most critical elements for conveying learning at the training institutes. One faculty member stated, "*The learning is in the doing*." Another faculty member stated, "*Tell me something, I learn a little bit; show me something and I'll learn a little more; let me do it and I'll learn the most*." Faculty discussed the many ways in which the participant projects enhance the learning process, including:

 Projects increased participant motivation to learn because they could see how the material could fit their own topics and help them to solve their own problems,

- Projects provided not only a sense of accomplishment, but an actual product they could walk out with and use once they returned home, and
- Projects gave trainees a focus for their efforts and allowed them to tie together the entire training institute's content as they learned.

Overall, faculty indicated that the training institutes would not be as effective without the use of individual projects. The process of working on individual projects usually involved working as part of a small group, and in some ways, learning to do this productively and efficiently was as important as completing a final product.

4.1.4 Blending of Instructional Methods

NIH Program Officers and training institute faculty members unanimously emphasized the importance of offering an *active learning experience* in which more time is spent on small group peer interaction and less time on didactics. Several respondents (both federal stakeholders and faculty members) indicated they had altered the relative emphasis of lectures versus group work to emphasize the latter, and that their trainee reaction questionnaire data supported this approach as being more valuable for the trainees.

Respondents from two training institutes raised a question about the choice of overall training institute orientation. The respondents identified two general approaches. In the first, there is an emphasis on completing an end product (i.e., a grant application or a well-developed concept that could be turned into a grant application)—the "end product" approach. One respondent said that organizing the training around the various stages of completing an end product was very well received by the trainees. However, a second respondent at a different training institute said they preferred a different organizational model based more on teaching methods and improving overall research capacity with a little grantsmanship added at the end—the "methods and capacity" approach. Both respondents agreed that the "end-product" approach was more popular with the trainees, but the second respondent preferred the "methods and capacity" approach because it was more intensive and demanding, and potentially more enriching for the trainees.

4.1.5 The Value of a Residential Immersion Learning Experience versus Online Training

NIH Program Officers and training institute faculty respondents agreed on the critical importance and value of the use of a residential setting for the training institutes. Respondents indicated that interaction with other trainees and with faculty is a critical component because it includes:

- Discussions and questions/answers after lectures, which allows for the emergence of new and unanticipated viewpoints and opinions;
- Informal discussion and contact as essential parts of the learning process;
- Interactions that build new relationships and community with other trainees and with faculty;
- Informal interactions that facilitate the development of persisting mentor-mentee relationships; and

• Germination of new ideas, insights, and new relationships.

Respondents also reported that the residential component allows trainees to be free from the usual demands of work, family, and home life, and this provides greater opportunity for deeper reflection and thinking. Respondents commented that immersion in the learning experience leads to greater intensity of the learning experience as evidenced by:

- Opportunities for informal interactions over meals, in evenings, in hallways, while exercising, and other leisure that builds relationships;
- Having the training, lodging and meals take place at the same facility (i.e., hotel, conference center) versus on campuses where people can go off on their own; and
- Trainees having time in the evenings and outside of the structured day to work on their concept papers and individual projects, which they may not have had if they were not in a residential setting.

While most of the NIH Program Officers and training institute faculty believe that offering the training in an online format would be less expensive and allow for greater reach to more trainees, their views about what would be lost from the training institute experience if it were provided online were equally as strong. Respondents argued that the biggest loss would be the opportunities for peer-to-peer and trainee-faculty interactions. Conducting the trainings online would make it more difficult to achieve the same level of trainee engagement and would reduce the capacity to develop new relationships and build community interactions. The loss of trainee engagement would be especially damaging to efforts to teach trainees how to work in multidisciplinary research teams. Other losses seen as likely from delivering the training online included:

- The experience would be much less intensive and not immersive;
- There would be technology limitations (not everyone has Skype or other instructional technology);
- It would seriously undercut the mentor relationships that have been an important part of the training institutes; and
- Participants would lose the "learning to think together" aspect of the training institutes.

These objections did not mean that respondents were completely opposed to conducting some aspects of the training institutes' content in an online format. Most respondents acknowledged that it is feasible for some types of content to be placed online, although they emphasized that there were caveats and important changes that would need to be made to the programs if this were to be done. Some of the caveats and possible changes included:

 One possibility (ATI) is to establish three tiers: one would be a basic online course available to any one; the second tier would be a more advanced presentation where each online session is facilitated by a live faculty member to allow for questions and answers and discussion; and the third would be away from the website and in a residential setting.
- Other training institutes suggested having both an online component and a residential component (ISSH) but emphasized that a really important part of the training is the formation of relationships which is difficult in a virtual setting.
- mHealth uses Google Hangouts, which allow for use of video (and might at least ensure that trainees are not answering email or other activities during the training session).
- Some individuals are better suited to learning in this way than others—most people seem to prefer the classroom setting.
- TIDIRH posts monthly online one-hour cyber-seminars on various dissemination and implementation research-related topics presented by a leading expert in a specific area identified as of interest to the trainees. There is opportunity during the live seminars to ask questions and hold some discussions, and questions can be emailed to presenters at a later time.

4.2 Recruiting and Selecting Faculty, Host Institutions, and Trainees

This subsection describes the second set of implementation processes—recruiting and selecting faculty, host institutions, and trainees.

4.2.1 Recruitment and Selection of Faculty

NIH Program Officers stated that the faculty recruited for the first years of the training institutes was selected from among the leading research investigators in each training institutes' scientific fields. Frequently, NIH Program Officers were personally acquainted with the investigators or could easily identify them from their grants and publications. As time passed, faculty members recommended additional candidates based on their expertise in newer areas. Respondents from each institute indicated there is usually a core faculty group that continues to teach over time, and guest faculty who teach on specific topics (some guest faculty may become core faculty). Recruiting faculty for the training institutes has not been a problem—word is out about the training institutes and prospective faculty members often contact NIH staff to volunteer. Where there has sometimes been a problem is with the availability of faculty—someone may not be able to participate during the time period selected for the training institute.

Interview respondents (both federal Program Officers and faculty) had very similar views about the prospective benefits or outcomes that faculty might experience from their participation in the training institutes. Involvement with the training institutes is considered prestigious by fellow faculty and by department chairs at academic institutions. Learning about their fields is commonly mentioned—faculty members get a lot out of the opportunity to discuss issues and problems with their colleagues and their trainees during the training institutes. They also learn about new examples or developments they can take back to their teaching at their respective academic institutions. New collaborations are formed—faculty may choose to collaborate with former trainees or with other faculty they meet at the institute. The chance to give something back is important—respondents stated that faculty enjoy mentoring their younger colleagues and find this rewarding in and of itself. The opportunity to enjoy a working vacation is a benefit—the training institutes for recreation. For training institutes

addressing emerging fields, the faculty gain recognition. One faculty member stated that it can be difficult to evaluate good work in emerging fields because department chairs often don't know how to measure success in emerging fields. Participating as a faculty member in an NIH training institute is one measure of success and recognition.

One problem mentioned is ensuring that faculty members (especially guest faculty members) stay for long enough to participate fully in the training institute. One respondent mentioned a faculty member at an early training institute who flew in from out of town, gave a 15 minute presentation and left immediately afterward. Some training institutes have imposed rules that require faculty members to stay for the entire training institute (or at least for several days) in order to allow for further interaction with trainees and other faculty.

Retention of faculty has generally not been a problem in that faculty members tend to be highly committed to participation in the training institutes. One training institute Program Officer mentioned working closely with faculty in advance of the training institute—reviewing the materials and slides they propose to use, even training new faculty members on mentoring.

4.2.2 Selection of Host Institutions

Four of the seven SRTIs rotate their training institutes among different host universities each year as shown in **Exhibit 13**. This is done to highlight the existing research program at the host university and provides the local research program with several important benefits: increased visibility within their local academic institution; publicity for the research program inside and outside the institution; and the opportunity to attract potential faculty in that research area in the future. Of the four training institutes that stage their training at different host institutions, three are emerging fields (mHealth, ISSH, and TIDIRH) while the fourth is the SW training institute.

I raining Institute	Host Institutions
ATI	 Same location each year since second year (Fluno Business Center, University of Wisconsin) Centrally located
C&AD	Same location each year (Bolger Conferencing Center, Potomac, Maryland)
ISSH	 Rotates among universities each year Intended to highlight the work the host institution is doing in a specific area of systems science Intended to reinforce institution's commitment to continuing work in system science Institutions vie for opportunity to host the training institute—considered very prestigious
mHealth	 Rotates among universities each year Considered good publicity for university, heightens visibility to NIH, attracts potential faculty
RCT	 Same location each year (Airlie Center, Warrenton, Virginia) Location is accessible for NIH program staff

Exhibit 13. Host Institutions

Training Institute	Host Institutions
SW	 Rotates among universities each year Selection of university is based on who is chairing the institute for that year Visibility to NIH
TIDIRH	 Rotates among universities each year Host institutions are allotted a few additional slots among trainees for that year Intention to highlight host institution's current activities and strengthen their capacity for D&I Some effort to ensure geographic variability—East one year, West one year, Central one year Host institution often supplies several guest faculty

The remaining three training institutes (ATI, C&AD, and RCT) use the same training/conference site each year and address established scientific fields.

4.2.3 Recruitment and Selection of Trainees

This subsection discusses the target populations identified by each SRTI, the selection process, and the availability of institutional work environment support for trainees after they return from the training.

4.2.3.1 Target Populations

Career stage is arguably the most important variable in defining the target populations for each of the SRTIs. The four training institutes that address established fields focus on new or early-stage research investigators as shown in **Exhibit 14**.

Training Institute	Target Population Career Stage	Established or Emerging Research Field?
ATI	New or early career	Established
C&AD	New or early career	Established
ISSH	All career stages (early, mid, senior)	Emerging
mHealth	All career stages (early, mid, senior)	Emerging
RCT	New or early career	Established
SW	New or early career	Established
TIDIRH	Post-doctoral early stage investigators with an interest in D&I research or investigators who have not yet been a PI on a D&I science grant.	Emerging

Exhibit 14	Target	Population	by	Career	Stage
------------	--------	------------	----	--------	-------

Some respondents mentioned other variables that were taken into consideration regarding the target population as shown in **Exhibit 15**.

Training Institute	Other Relevant Variables/Considerations	Established or Emerging Research Field
ATI	Although sponsored by NCI, they wanted all types of disciplines	Established
C&AD	Researchers from different disciplines who were broadening their methodological approaches such as how to work with schools and hospitals	Established
ISSH	This institute sought to promote racial and geographic diversity in its trainees	Emerging
mHealth	In the first year, they focused on junior investigators; however they opened admission/selection up in subsequent years to get word out to the broader research community (grow the field))	Emerging
RCT	Focused on junior researchers because they wanted trainees who would be more likely to make a commitment to behavioral clinical trial research	Established
SW	Selection emphasis on Social Work schools at Tier 2 research level (Carnegie) because these are schools that are interested in breaking into research	Established
TIDIRH	More experienced investigators do not receive mentoring—for them there would be more emphasis on peer networking	Emerging

Exhibit 15. Other Relevant Target Population Variables/Considerations

4.2.3.2 Trainee Selection Process

The trainee recruitment and selection process is generally similar for each of the SRTIs. Applicants can learn about the availability of a training institute on the OBSSR, NCI, or NICHD websites. They also learn about the training institutes through listservs, conferences, former trainees at their institutions, and other faculty. ISSH appears unique in that its main recruitment tool was its listserv, although information about each year's training institute can be found on the OBSSR website. As shown in **Exhibit 16**, applicants are requested to provide several types of materials in addition to their application.

Training Institute	Description of Participant Selection Materials	Established or Emerging Research Field
ATI	Letter of recommendation, CV, personal statement (including research interests and how the institute would further those), and a list of completed graduate courses in Behavioral Sciences and Statistics	Established
C&AD	Personal statement of career and research interests (including a description of specific applied research topic; bio-sketch, two letters of recommendation (one on applicant's qualifications and the other describing the applicant's professional environment including research support and resources)	Established
ISSH	Description of a proposed project relevant to the track to which applicant is applying, a CV, letters of recommendation.	Emerging

Exhibit 16. Participant Selection Materials

Training Institute	Description of Participant Selection Materials	Established or Emerging Research Field
mHealth	Summary of professional background, research interests (including a specific testable research question relevant to mHealth methods; description of a research team experience; CV, and letters of support	Emerging
RCT	Institution information, sponsor/reference letter, CV, personal statement describing interest in the institute, training goals, and how the training is beneficial to pursuing those goals	Established
SW	Two-page concept paper proposing an intervention research project, CV, letter of support from Dean or Department Chair, one- page personal statement describing research and career interests and how the training will benefit them	Established
TIDIRH	Statement of interest, current CV, a research concept paper, and recommendations from senior researchers at their institution, no more than two trainees from the same institute can be accepted for the same cohort	Emerging

Materials from applicants are reviewed by a selection committee from each training institute. The committees establish a set of rating criteria against which each applicant's background and additional statements are evaluated. Criteria might include: the Personal Statement, Research Statement, previous experience with content area, CV, and letters of recommendation. Each criterion might receive a different weight, and the final score, based on summing these weights, would need to exceed a predetermined threshold to result in selection for the training institute for that year.

While this approach sounds precise, interview respondents noted that it was not without problems. The specific criteria might change from one year to the next, the relative importance of different criteria might change in a similar fashion, and overall there was a trend with passing time that the threshold for selection might increase as more individuals applied for the same number of available slots. In addition, the individuals conducting the selection process might change over time. Thus, a trainee who was selected during an earlier year of a training institute might not qualify for selection during a later year.

4.2.3.3 Availability and Quality of Data on Past Years' Trainees and Unselected Applicants

Evaluation Advisory Committee (EAC) members were asked to provide the evaluation team with data on past trainees and unselected applicants as part of the document review process⁵. This was important information for constructing possible comparison groups. We reviewed this information to assess the extent to which the data contained:

- Adequate coverage for each year the training institute was offered;
- Contact information on selected trainees and unselected applicants;

⁵ At a meeting of the Evaluation Advisory Committee on August 18, 2014, some EAC members stated that it was possible that some additional data on selection criteria and scoring may exist that had not been provided to Madrillon.

- Indications of what criteria were applied in the selection process; and
- Indications of how these criteria were weighted to produce a selection decision.

Review of the data found serious limitations on each of these four parameters. All of the seven training institutes were able to report the names (and usually the contact information) for the trainees selected for each year the training institutes were offered. The availability of similar information for unselected applicants was far less consistent. Types of information reported (using Excel spreadsheets) included name, address, highest degree, institution (or company), job title, discipline, and (less often) statements on background and research interests. These types of information were not present for all years for many of the training institutes. Only two training institutes provided information on the types of criteria used in the selection process and how these were combined. Thus, for most of the training institutes, the most consistent information included alphabetical listings of selected trainees and unselected applicants. A few training institutes also identified a small number (5-10 individuals) of wait-listed applicants. While these would be more comparable to the selected trainees, the numbers of wait-listed applicants were too small to form an adequate comparison.

4.2.3.4 Institutional Post-Training Workplace Supports

Six of the seven training institutes require that applicants submit letters of recommendation from more senior individuals within the applicant's institution (e.g., department chairs, deans, etc.). In addition to providing personal endorsements of the trainees' research skills and potential, these letters also describe the types of supports the institutions will provide to trainees to promote their research careers when they return from the training. The training evaluation literature has shown convincingly that the presence or absence of adequate workplace support can make or break the trainee's success in applying what has been learned. The letters of recommendation are therefore carefully reviewed as part of the trainee selection process for each training institute, and if there are no indications that an institution will support the trainee beyond providing release time to attend the training, that applicant is not selected.

Some of the training institutes attempt to address this issue by including time during the training to assist trainees in formulating a plan for how they will use what they have learned to advance their research careers. The mHealth training institute includes a specific session on professional development on the trainees' last day of training. The types of new partnerships and collaborations that trainees need to develop may be very different from those more commonly found in their home research institution. The training institute developers were aware of the danger that trainees could end up as "stuck between two worlds." Trainees are therefore prompted to reflect on how they can pursue their new research interests in mHealth research in a more traditionally oriented research setting in ways that do not marginalize them.

Results from the training institute faculty interviews indicated that institutional supports can take a number of forms. Providing coverage to permit the trainee to attend the training institute was one type of support. This required providing release time from any teaching or clinical responsibilities, and for some training institutes, providing the cost of travel to the training site. Supports for trainees after their return from training

proved more variable. Support can take a variety of forms: support from senior faculty to pursue grant application and publication writing, reduced teaching or clinical loads, provision of pilot research funding, access to consultation with statistical and data analysis staff and use of research assistants, mentoring by locally well-established research investigators, and access to lab facilities and datasets that would allow trainees to gather pilot data.

4.3 Key Findings on Structure and Implementation

This discussion of the structure and implementation process for the SRTIs has emphasized some important similarities and differences across the training institutes. These include:

- The design of the SRTIs incorporated principles of adult learning theory. Adult learning theory characterizes adult learners as internally motivated individuals who learn best when the content is perceived as relevant to their personal needs and is presented using a mix of instructional methods. According to adult learning theory, when these conditions are met, adult learners are more engaged and involved in the learning experience, thereby increasing their ability to absorb new knowledge and learn new skills.
- While the specific curriculum developed for each training institute differed in content to address the respective scientific fields and methods, each curriculum shared some common elements (for example, basic theories and methods).
 Four institutes included material on grantsmanship, and three included instruction on developing multidisciplinary research teams and building new research fields.
- All seven training institutes updated their course content over time based on two factors: new developments or controversies within the respective scientific fields, and feedback from faculty members and trainees who completed the training. Two training institutes organized their training around different themes each year, but a growing trend across training institutes was the increasing importance of designing and evaluating behavioral interventions.
- Three other components of the training institutes were identified by respondents as critical to the success of the institutes. Mentoring was a formal element of five of seven training institutes. Mentors were typically assigned to work with individual trainees or small groups based on common research interests. While the training institutes encouraged trainees to maintain contact with their mentors after the training, faculty members reported that only some did so. Networking was an informal element of the training institutes and was greatly facilitated by the residential setting. Four training institutes established listservs or used social media to promote ongoing contact and communication among trainees after the training ended. The success of these listservs depends upon the ongoing commitment of NIH Program Officers involved with each training institute; one training institute maintained a listserv for six years until the Program Officer who maintained it left for a different position. Individual (or group) projects were the third critical element of the training institutes. The use of these projects as teaching tools enabled many trainees to complete their training experience with a finished product (a grant proposal or a solid outline

for one) and a process that enabled them to integrate their new knowledge and skills. In the training institutes that used group projects, trainees could learn additional skills in working with a multidisciplinary team.

- Faculty recruitment was initially based on each training institute's planning group's contacts with research investigators who were leaders in their fields. Once each training institute had completed one or two years of training, faculty were able to identify additional investigators who were expert in newly emerging topics. Each training institute had a group of core faculty who taught over several years, and guest faculty who would teach on a particular topic one year (and might subsequently become part of the core group.) While there was turnover among faculty, it was not difficult to retain qualified faculty members, in part because prospective faculty members viewed the training institutes as prestigious and enjoyable opportunities.
- Four training institutes rotated the site of each year's institute among universities that had active research programs in that field. Selection as a host site was a prestigious honor for a local research group or program, because it strengthened ties with NIH, increased the visibility of the research unit or group within their own institution, and provided an informal mechanism for recruiting potential new faculty members.
- The SRTIs targeted potential trainees largely by career stage. The four training
 institutes addressing established scientific fields targeted new or early stage
 research investigators in order to train individuals who they hoped would
 commit to a research career in that field. The three emerging-field training
 institutes tended to seek both early stage and more experienced investigators in
 order to encourage a more rapid development of research portfolios in these
 fields. An important aspect of the emerging-field training institutes was that they
 often brought together trainees that had not typically worked together in the
 past—for example, ISSH attracted engineers and information and computer
 scientists as well as health and biomedical researchers, and mHealth brought in
 people from industry as well as researchers.
- Federal and faculty interviewees emphasized that the trainees selected to attend the seven training institutes represented a highly selected group comprising individuals who had already demonstrated considerable promise and some success in their research careers. The trainee selection process was highly competitive; in some cases, more than two hundred applicants applied for 30 or 35 slots each year. As a result, the training institutes' selection committees had a surfeit of highly qualified individuals from which to select.
- The selection process was similar in general outline. Candidates submitted an application package consisting of a curriculum vita, a personal statement describing research and professional goals, a research statement outlining an idea or possible research project that would be worked on during the training, one or more letters of recommendation, and, in some cases, a list of graduate courses completed in the scientific field. A trainee selection committee that included NIH Program Officers and sometimes outside faculty would establish a set of criteria based on these materials, and committee members reviewed and rated the applicants' materials based on these criteria. The ratings were

summed to create an overall score, and this score would be compared against a threshold, with those applicants scoring above the threshold being accepted and those scoring below rejected. Rejected applicants could (and sometimes did) reapply in subsequent years.

- Interviewees expressed several problems with the selection process. Criteria, committee members, and the selection threshold, might change from one year to the next. Interviewees noted that in the later years of each training institute, the selection threshold did change such that individuals accepted in an earlier year might not have been accepted in a subsequent year. Recordkeeping on the rating criteria, scores, and thresholds was inconsistent and substantially lacking in earlier years of the older training institutes.
- The available data on unselected applicants suffer from serious limitations. Data were not available for all unselected applicants for all years each training institute was offered, and usually lacked information on what criteria were used to select trainees and how these criteria were combined to reach a selection decision. While a few institutes did identify applicants who were placed on a waiting list (and therefore might have been more comparable to the selected trainees), the numbers of wait-listed individuals did not exceed ten in any one year. This was too small to form an adequate comparison group.

V. Outcomes from the Summer Research Training Institutes

This section sets the stage for the development of an evaluation design for the trans-NIH Summer Research Training Institutes. The analyses of process variables in the previous section showed that the seven training institutes share some important similarities as well as a number of important differences. The following discussion considers two topics that are critical for the evaluation design: expected outcomes, and the feasibility of comparison conditions.

In thinking about the possible effects these training experiences could have, we make an important assumption: the SRTIs do not *cause* outcomes directly, but they do *contribute* (as one factor among others) to various post-training results. Many arguments can be made to support this assumption: the SRTIs train participants over a short interval of time (between 5 to 10 days); the trainees chosen by the selection committees to attend the training institutes represent individuals who were already on the verge of success as research investigators; and success in a research career is influenced by the specific research milieu to which the trainees return are only a few of many. Given the assumption that training contributes to rather than causes change, it clearly becomes important to understand not only what kinds of effects the training is expected to produce, but how and why these effects occur.

There are three approaches we can take to examine these effects and their rationale. Theoretical models and frameworks on training evaluation provide a useful starting point because they suggest the kinds of outcomes that are important to consider (and often furnish some ideas concerning why these outcomes should occur). Therefore, as part of the evaluation team's literature review for the Feasibility Study, we reviewed theoretical models and frameworks currently popular in the evaluation training literature. Another avenue by which to identify potential outcomes is through review of past evaluations of similar programs. The second part of our literature review examined a total of 18 publications describing 21 evaluation studies of NIH-sponsored summer research training programs. We report those results here. The final approach was to ask the NIH Program Officers and training institute faculty members what types of outcomes they expected to see from the SRTIs.

5.1 Theoretical Models and Frameworks in the Training Evaluation Literature

5.1.1 The Kirkpatrick Four Levels® Model

The dominant evaluation model used in training evaluation today is the Kirkpatrick Four Levels® Model. The Kirkpatrick® Model grew out of Donald L. Kirkpatrick's doctoral dissertation entitled *Evaluating Human Relations Programs for Industrial Foremen and Supervisors* and subsequent articles and books published over the next 45 years. The model provides guidance to individuals charged with conducting evaluations of industrial and corporate training programs by identifying four levels at which training evaluation should occur. These four levels include assessment of trainees' *reactions* to the training experience, trainee *learning* (new skills and knowledge gained as a direct result of the training), trainee *behavior* (trainee application of the new knowledge and skills when he or she returns to the work site), and *organizational results* (the benefits from the training to the organization).

The Kirkpatrick® Model was first introduced through a series of four articles in the journal of the American Society or Training and Development (ASTD) in 1959 and 1960. Following the appearance of these articles, the Kirkpatrick® Model quickly became the standard conceptual framework used in training evaluations (McLean and Moss, 2003). In the late 1990s, the CDC adopted the Kirkpatrick framework as a core element in its Training Effectiveness and Efficiency Model (TEEM-Dunet and Reyes, 2006) and in 2011, the United States Office of Personnel Management's *Training and Field Guide—Demonstrating the Value of Training at Every Level* (US Office of Personnel Management, 2011) used the latest iteration of the Kirkpatrick framework as the basis for teaching managers and training specialists how to plan and conduct training evaluations.

In their 2009 white paper entitled *The Kirkpatrick Four Levels*[™]: A Fresh Look After 50 Years, Jim Kirkpatrick (Donald's son) and Wendy Kayser Kirkpatrick note that the model incorporates both a planning process for building effective training programs as well as the evaluation process as shown in **Exhibit 17**). In his 1993 book of case studies of evaluations using the framework, Donald Kirkpatrick (1993; p.26) explained that "trainers must begin with desired results and then determine what *behavior* is needed to accomplish them. Then trainers must determine the attitudes, knowledge, and skills that are necessary to bring about the desired behavior(s). The final challenge is to present the training program in a way that enables the participants not only to learn what they need to know but also to react favorably to the program." Thus, planning starts with Level 4 and moves down through the levels, whereas data collection for the evaluation begins at Level 1 and moves upward.

Exhibit 17. The Kirkpatrick Four Levels® Model as a Planning and Evaluation Approach PROGRAM & EVALUATION PLANNING Level 4: RESULTS The degree to which targeted outcomes occur, as a result of the learning event(s) and subsequent reinforcement.			
		The degree to which targeted outcomes occur, as a result of the learning event(s) and subsequent reinforcement.	Î
	Level 3: BEHAVIOR	The degree to which participants apply what they learned during training, when they are back on the job.	
	Level 2: LEARNING	The degree to which participants acquire the intended knowledge, skills, and attitudes based on their participation in the learning event.	
Ļ	Level 1: REACTION	The degree to which participants react favorably to the learning event.	DATA COLLECTION

Kirkpatrick and Kirkpatrick (2009) emphasize the importance of engaging stakeholders across the organization in planning for a training event and its evaluation. They note (p.4) that "contrary to training myth and deep tradition, we do not believe that training events *in and of themselves* deliver positive, bottom line outcomes. Much has to happen *before* and *after* formal training. Learning professionals need help to do it, and

that help has to come from our business partners." The current Kirkpatrick process emphasizes careful attention to the post-training work environment, where, according to a statistic they cite, 70% of actual learning takes place. They strongly emphasize the role of the supervisor or manager in reinforcing newly acquired knowledge and skills.

The evaluation philosophy that the Kirkpatricks bring to training evaluation is an important part of the Kirkpatrick Four Levels® model. Beginning with an assessment of the expectations various stakeholders have for training impacts (that is, what outcomes do they intend for the training to achieve), their approach involves constructing a *chain of evidence* that shows that these expectations were clearly met. The chain of evidence includes both quantitative and qualitative data that can demonstrate convincingly that the training has *contributed* to achievement of the expectations of the stakeholders. To be convincing, it is necessary that the evidence assembled demonstrates not only that the impact was achieved, but how it was achieved. This approach sidesteps the usual evaluation quest to demonstrate attribution through comparison against a control group, and takes the position that constructing a convincing chain of evidence can suffice to demonstrate a meaningful return on expectations. In its logic, this approach is similar (although not identical) to contribution analysis (Mayne, 2008; Lemire, 2010).

5.1.2 Alternatives to the Kirkpatrick Four Levels® Model

While the Kirkpatrick Four Level Models® dominates training evaluation practice today, it is not the only framework that has been suggested for evaluating training initiatives. Reactions to the Kirkpatrick® Model have followed two paths. The first path consists of researchers who have modified the Kirkpatrick® Model by adding new levels, or splitting current levels, or both. An example of the latter is the Return-on-Investment framework developed by Jack Phillips (1997). His approach seeks to isolate the effect of the training from other environmental factors through the use of a control group. His five-level model includes reaction and intended action when trainees return to work (Level 1), learning (Level 2), job application (Level 3), business results (Level 4) and a new return on investment to the organization (Level 5). Kaufman, Keller, and Watkins (1995), and Watkins et al. (1998) proposed what they termed the Organizational Elements Model, which introduced consideration of resources and processes as important parts of the model. They divided Kirkpatrick's reaction level into resource availability and guality, and process acceptability and efficiency. They retained Levels 2-4 and added a new fifth level (Societal Contributions). Wang and Wilcox (2006) reframed the Kirkpatrick® Model by focusing on short and long-term outcomes; reactions and learning are considered short-term outcomes while behavior on the job, organizational impacts, and return on investment are long-term outcomes.

Nikols (2005) discussed a different perspective that he argues is compatible with Kirkpatrick. Nikols' Stakeholder Model states that training must satisfy the needs and interests of multiple stakeholder groups, which he defines as groups or individuals that contribute something to the training process and in turn expect some type of benefit from it. These contributions and expectations of benefits may differ markedly across the various participating groups. To evaluate a training initiative properly, it is necessary to identify what each stakeholder group contributes to a training initiative, what benefits each group expects to receive from it, and the extent to which each group is satisfied

with what it has received. This approach is similar to the evaluation approach inherent in the Kirkpatrick® Model as described above. Another approach with some similarities to both of these is described by McLinden and Trochim (1998), who use multidimensional scaling as a means of quantifying stakeholder expectations and satisfaction with results.

A second type of reaction by critics of the Kirkpatrick® Model is evident in more recent theoretical frameworks by learning theorists who decry its lack of consideration of cognitive, affective and behavioral variables. Models including these variables have recently been introduced (Kraiger, Ford and Salas, 1993; Kraiger and Merritt, 2010; and Holton, 1996). There are some published studies in the training evaluation literature in which these models have been applied, but their complexity and the burden of measuring numerous variables have worked against their widespread adoption by training evaluators.

A final approach that warrants discussion is Brinkerhoff's (2005) Success Case Method. This is a mixed methods evaluation approach that combines narrative case studies with quantitative methods through a five-step process. To apply the method, it is first necessary to develop an impact model of the training, identifying the goals of the training and specifying what success looks like in light of these goals. Then, the evaluator conducts a survey of training participants to identify "best" and "worst" cases. The evaluator then selects a sample of best and worst trainees and obtains corroborating evidence from these trainees; this evidence usually derives from interviews but can include document review or other data that support the classification. The evidence focuses on how the trainees applied or were unable to apply the knowledge, skills, and attitudes learned in the training. The resulting data are analyzed to identify success factors, barriers, and challenges in transferring the learning into the worksite. Additional data can be obtained from other sources such as supervisors. The Success Case Method can be used as a stand-alone evaluation approach or combined with another evaluation approach.

5.2 Review of Earlier Evaluations of NIH-Sponsored Summer Research Training Institutes

Using a rigorous definition of NIH-sponsored summer research training institutes, the evaluation team identified 18 published evaluation reports that dated from 1997 through 2013 and are shown in **Exhibit 18**. The 18 publications reported on a total of 12 distinct summer research training institutes. Six of the 12 training institutes included sponsors from outside NIH, including the VA, CDC, the International Atomic Energy Association—Programme of Action for Cancer Treatment (IAEA-PACT), the Health Research Board (HRB) and the Health and Social Care Research and Development Division of Ireland (HSCR&D), the American Association for Geriatric Psychiatry (AAGP), the John A. Hartford Foundation, and the Janssen Pharmaceutical Foundation. There were two or more evaluation publications for 4 of the 12 training institutes. Three of the publications (Waitzkin et al., 2006; DeRouen and Wiesenbach, 2012; and Mehrotra et al., 2013) reported results for two separate evaluation samples, producing a total of 21 evaluation samples for the 18 articles.

Exhibit 18. Summer Research Training Institutes, Sponsoring Institutes, and Evaluation Study Citations

Name of Summer Research Training Institute	Sponsor	Evaluation study
Advanced Research Institute in Geriatric Mental Health	• NIMH	Bruce et al. (2011)
Implementation Research Institute	NIMH	Proctor et al. (2013)
	• VA	Stamatakis et al. (2013)
Minority Research Infrastructure Program (M-RISP) & Mentorship and Education Program (MEP)	• NIMH	Waitzkin et al. (2006)
Research Career Development Institute for Psychiatry	• NIMH	Kupfer et al. (2009)
Principles and Practice of Cancer Prevention and	NCI	Faupel-Badger et al. (2011)
Control Annual Summer Course	IAEA-PACT HBB	Otero et al. (2012)
	HSCR&D	Williams et al. (2013)
Summer Genetics Institute	NINR	Hickey et al. (2013)
Summer Institute in Clinical Dental Research Methods	 NIDCR Fogarty International Center 	DeRouen and Wiesenbach (2012)
Summer Research Career Development Institute	NCMHD	Berget et al. (2010)
Summer Research Institute in Geriatric Psychiatry	NIMH	Halpain et al. (1997)
	AAGP VA	Halpain et al. (2001)
	 Janssen Pharmaceuti cal Foundation 	Jeste et al. (2007)
Summer Research Training Institute for American Indian and Alaska Native Health Professionals	NIGMSIHSCDC	Zaback et al. (2010)
Summer Training Program in Aging Research for	• NIA	Mehrotra et al. (2009)
Social Workers	 OBSSR John A. Hartford Foundation 	Mehrotra et al. (2013)
Training Institute for Dissemination and Implementation Research in Health	 OBSSR NCI NIMH VA 	Meissner et al. (2013)

5.2.1 Characteristics of the Summer Research Training Institutes

The 12 summer research training institutes shown in **Exhibit 18** addressed four main objectives including:

1. Assisting new or early-stage investigators in developing the grantsmanship skills needed to obtain their first R01-level grants;

- 2. Providing career development assistance to minority researchers;
- 3. Building research capacity in under-developed research fields; and
- 4. Training researchers and health professionals on specialized topics.

Five of the 12 summer training institutes explicitly target new or early-stage investigators and 4 target minority investigators. Other training institutes seek to build research capacity in fields where more research investigators are needed (e.g., the Summer Institute in Clinical Dental Research Methods, and the Summer Training Program in Aging Research for Social Workers) while some institutes focus on providing training on specialized topics (e.g., the Summer Genetics Institute, and the Principles and Practice of Cancer Prevention and Control Annual Summer Course). These objectives shape the structure of these training initiatives. Those initiatives that intend to support new or early-stage investigators or minority researchers tend to include an active mentoring component that can continue for as much as two years after the initial summer research training institute. In addition, some summer institutes offer additional booster sessions or provide for continued interaction among trainees after the summer institute through specialized Internet websites. Thus the activities of summer research training institutes may continue long past the single summer experience, emphasizing the need to incorporate process characteristics such as how a training intervention is structured as part of a larger training initiative as an element of the evaluation design. A final structural element is the provision of limited funding in some of the training initiatives to support pilot data collection or travel expenses for visits to successful research projects. This element would need to be considered in any cost analysis.

To be eligible for selection for these training institutes, applicants generally were required to demonstrate potential for a successful research career. For some training institutes, this meant submission of a personal statement of commitment to a research career in the field addressed by the institute (e.g., the Summer Research Institute in Geriatric Psychiatry, the Implementation Research Institute, and the Training Institute for Dissemination and Implementation Research for Social Workers, required submission of an abstract, a concept paper, or a draft research proposal. The Advanced Research Institute in Geriatric Mental Health restricted eligibility to Early-Stage Investigators who had obtained an NIMH K-award, while the Training Institute for Dissemination and Implementation Research Institute for Dissemination and Implementation Research proposal. The Advanced Research Institute in Geriatric Mental Health restricted eligibility to Early-Stage Investigators who had obtained an NIMH K-award, while the Training Institute for Dissemination and Implementation Research Institute and the Research Career Development Institute for Psychiatry required written evidence of a commitment from a local mentor with a demonstrated research grant history.

The summer research training institute components were offered for different periods of time, ranging from 2.5 to 3.5 days, to as long as 8 weeks. Some also offered additional booster sessions, sometimes tied in with meetings of national professional associations that trainees were likely to attend.

5.2.2 Characteristics of Summer Research Training Institute Evaluation Designs

Exhibit 19 shows the percentage of the 21 evaluation studies using various outcome measures. The two most consistently used outcome measures were: 1) new grants primarily addressing the content area of the training institute, and 2) the number of publications, including those addressing the content area of the training institute, in peerreviewed journals where the trainee was the lead author or reported as a co-author. Nineteen studies used the number of new grants as an outcome measure of the training institute. One evaluation analyzed whether trainees were PIs of the grants received. Award measures were collected through surveys as self-report, searches of NIH grant databases, and/or ongoing communications by faculty from the institute or with the mentors of participants who kept track of progress being made by trainees after the training. Also documented by nearly half of the evaluation studies was the number of grant applications submitted to NIH. Mehrotra et al. (2013) reported that trainees who had

Exhibit 19. Outcome Variables Used Ir	١
Summer Research Training Institute	
Evaluations	

Outcome Measure	% of Evaluation Studies Using It
New grant awards in training institute field	90%
New publications in training institute field	81%
Career advancement	62%
Increased motivation to write grants	57%
Maintaining contact with mentor	38%
New grant submissions in training institute field	38%
Maintaining contact with other trainees	33%
Using training institute content in teaching	33%
Memberships in content- related professional associations	33%
Continued contact with NIH Institute	24%

participated in NIA's training on aging research in social work between the years 2004 and 2009 had secured a total of more than \$10 million in grant awards from federal and non-federal sources. This total included 26 grants of \$100,000 each from the John A. Hartford Foundation's Geriatric Faculty Scholars Program. Officials at the John A. Hartford Foundation were so impressed with this finding that they subsequently chose to participate in and partially fund the NIA summer research training initiative in subsequent years.

The evaluation designs reported in the 18 evaluation publications included four patterns defined by their use of end-ofinstitute (exit) and/or follow-up surveys (see **Exhibit 20**). The most common approach (n=7) involved collection of data when

Exhibit 20. Evaluation Designs Used in Summer Research Training Institute Evaluations

A+B: Exit + Follow-up at a Defined Interval	7
A+B: Exit + One-Time Follow-up	5
B Only: Follow-up Only	6
A Only: Exit Only	3

trainees completed the training with a follow-up survey at a well-defined interval (e.g., 6 months or 12 months) after the training. A second large group (n=6) conducted a follow-up survey of multiple cohorts. A third group of 5 studies reported exit data and results

from a one-time survey of multiple cohorts (but these data were not linked). The remaining studies reported exit data only.

Only 4 of the 21 samples involved the use of a comparison group, and these four evaluation articles illustrated three approaches for constructing a comparison group. The approaches included selection of a comparable group of individuals who had not applied for the training institute; selection of individuals from among applicants not accepted for the training institute; and within-group comparisons of participating trainees based on trainee characteristics.

In addition, all of the training institutes claimed success in meeting their goals; none of the evaluations reported the use of a logic model, conceptual framework, or program theory; most evaluation reports included measures that corresponded to some aspect of Kirkpatrick's® Level 1 (Reaction); and very few reports described any attempt to measure (formally or informally) whether learning had taken place (Kirkpatrick® Level 2). Many of the reports referenced use of outcome measures that reflect application of the training at their academic institutions (Level 3); the most frequently listed outcomes were grant awards in the institute's field and new publications in this field.

5.3 Outcomes Identified by NIH Program Officers and Trans-NIH Summer Research Training Institute Faculty Members

The interview protocols for the NIH Program Officers and the SRTI faculty members included identical questions asking respondents to identify what they believed to be the most important short-term (immediate), intermediate, and long-term outcomes for their respective training institutes. The multiple responses for each of these three temporal categories were reviewed and collected within coding categories for the seven institutes. The data presented in **Exhibit 21** show these categories for both groups of respondents. The NIH Program Officers are represented by the **red checkmarks** (\checkmark) while the training institute faculty are shown with **blue checkmarks** (\checkmark).

Exhibit 21 shows that there was a considerable diversity of opinion about the types of outcomes NIH Program Officers and institute faculty members expected. As a rough index of what respondents from both groups combined thought were the most important outcomes, we focused on those for which respondents from four or more of the seven training institutes identified and considered these by *immediate* versus *intermediate* versus *intermediate* versus *long-term*. These outcomes are highlighted in yellow in the Exhibit.

In terms of *immediate* outcomes, at least one respondent from all seven of the training institutes identified increased understanding of basic principles, theories, and methods of the scientific field taught at the training institute as an important expected outcome. At least one respondent from four of the training institutes identified increased motivation and self-efficacy to design and conduct research in the institute field as an important expected outcome. Three of the four respondents endorsing this were institute faculty. The same was true of the development and implementation of a personal plan for achieving independent research support.

TYPE OF	CODED CATEGORIES FOR FEDERAL	ATI	C&AD	ISSH	mHealth	RCT	SW	TIDIRH
OUTCOME	STAKEHOLDERS' OUTCOMES	Existing	Existing	Emerging	Emerging	Existing	Existing	Emerging
IMMEDIATE	Increased understanding of basic principles, theories, and methods of training institute topic	✓	$\checkmark\checkmark$	√ √	✓	✓	\checkmark	✓ ✓
IMMEDIATE	Increased ability to evaluate and appraise the use of principles, theories, and methods in others' publications	✓				✓		
IMMEDIATE	Increased motivation/self-efficacy to conduct research on training institute topic	\checkmark	\checkmark		✓		\checkmark	
IMMEDIATE	Increased awareness of additional materials and resources trainees can use to learn more about topic	✓		√ √		✓		~
IMMEDIATE	Increased ability to apply principles, theory, and methods to trainees' own research problems	✓	✓	√ √			\checkmark	
IMMEDIATE	Development and implementation of a personal plan for achieving independent research support	✓	✓	√ √			\checkmark	
IMMEDIATE	Interest in a specific research idea/problem	\checkmark	\checkmark				\checkmark	\checkmark
IMMEDIATE	Increased number of contacts (other trainees, faculty, NIH program officers)				✓			
INTERMEDIATE	Increased ability to apply principles, theory, and methods to trainees' own grant applications	✓	\checkmark	√ √		✓	√ √	✓
INTERMEDIATE	Conducting pilot studies using these methods			\checkmark				
INTERMEDIATE	Increased collaboration with others (especially other trainees and institute faculty)	✓	✓	√ √	√ √			✓
INTERMEDIATE	Increased research productivity—research publications	\checkmark	\checkmark	\checkmark	✓	\checkmark	\checkmark	<
INTERMEDIATE	Increased research productivity—grant applications		\checkmark	$\checkmark\checkmark$	✓	\checkmark	\checkmark	✓
INTERMEDIATE	Maintenance of ongoing contact with institute mentor(s)		\checkmark	✓			\checkmark	
INTERMEDIATE	Increased contact with NIH program officers (especially during grant application development)		✓	✓				
INTERMEDIATE	Inclusion of principles, theory, and methods in teaching and mentoring	√ √		✓				✓
INTERMEDIATE	Increased formation of and/or participation in multidisciplinary research teams related to institute topic			✓	√ √	✓		✓
LONG-TERM	Efforts to promote research activities/program development at home academic institution	✓		√ √				
LONG-TERM	Efforts to increase visibility of scientific field at NIH			✓	✓	\checkmark		
LONG-TERM	Efforts to increase visibility of scientific field through professional societies and associations				✓			
LONG-TERM	Increased career progression	\checkmark	\checkmark	\checkmark	✓		\checkmark	$\checkmark\checkmark$

Exhibit 21. Outcome Categories Identified by NIH Program Officers and SRTI Faculty

The most important expected intermediate outcomes were an increased ability to apply principles, theory, and methods to trainees' own grant applications, increased productivity in terms of field-relevant publications and grant applications. Increased collaboration with others (especially other trainees and faculty from the institute) was seen as important by at least one respondent from five of the training institutes, and increased formation of and/or participation in multidisciplinary research teams on projects related to the institute's field was important for respondents from four institutes.

The most important expected long-term outcome was increased career progression.

The outcomes identified by our interview respondents were similar to those that had been identified through the literature review, although the interviewee responses were more nuanced in terms of the immediate and intermediate outcomes.

5.4 Pilot Study of Trainee Satisfaction with the ISSH Training Institute

The evaluation team conducted a pilot study using all available training satisfaction forms from the four years of the ISSH training institute (2009-2012). The pilot study assessed the level of depth and detail available in training institute satisfaction forms and identified possible outcomes based on a content analysis of the narrative responses trainees made to several items. The ISSH offers three tracks (social network analysis, agent-based modeling, and system dynamics modeling. A total of 200 trainee satisfaction forms were available for the four years; 62 forms assessed the social network track, 80 forms assessed the agent-based modeling track, and 58 forms were available for the system dynamics track.

Based upon our analyses of trainees' comments on the 200 surveys, we found that the trainee satisfaction forms are an important source of data for identifying potential outcomes. There were four major findings from these analyses:

- Most trainees reported that they had two primary objectives in attending the training; these included obtaining hands-on experience with the specific method and accompanying software, and meeting and networking with others (both trainees and faculty) interested in applying these methods in health and biomedical research.
- Trainees were most interested in using the time spent in their respective tracks to apply the methods and software to real-world research questions and data, and were seeking advice and assistance in applying these tools to their own research.
- Nearly all trainees indicated that they had made contacts colleagues with whom they would like to collaborate on research activities in the future, but the system dynamics trainees felt least ready to begin those collaborations immediately.
- Trainees from all three tracks stated that the training had reinforced their interest in using the specific method in their personal research and were interested in learning more about how to do so.

The evaluation team noted that these immediate outcomes were similar to those identified by NIH Program Officers and training institute faculty members. This pilot study showed that the analysis of narrative comments from training satisfaction forms can provide useful information on Kirkpatrick® Level 1 and 2 outcomes.

5.5 Perspectives on the Use of Comparison Groups

5.5.1 Perspectives of NIH Program Officers and Training Institute Faculty Members

The evaluation team was especially interested in exploring the perspectives of NIH Program Officers and training institute faculty members on the feasibility of a comparison group and how a comparison could be constructed. Several themes characterized their responses to these questions. The NIH Program Officers were split on the question of whether a comparison group was necessary. About one-third of these respondents stated that they did not think a comparison group was necessary at this time. One respondent expressed this viewpoint: "I don't know that a comparison group would be especially helpful at this point. I think knowing more about the participants is important." The remaining NIH Program Officers indicated that they were not sure, or did not know whether a comparison group was needed. This uncertainty was associated with skepticism that a good comparison could be found. Typical of this perspective was the following respondent's comment: "You could do a straight outcome study looking at what the participants achieved and feel good about that, but to what extent are those achievements attributable to our intervention? What would they have accomplished without our training? We take people who are right at the line for success and give them a push—would they have been successful anyway? I don't know." Another respondent noted that "we can't say that because of the institute people got this or that award, but if not for the institute they probably would not have competed as effectively...You have to consider questions of attribution and contribution. Can we attribute success to the training or can we say it made a contribution to their success? These are not the same thing. We picked people on a successful trajectory, and the fact that they followed it may have happened anyway."

When asked to discuss possible types of comparisons, the Program Officers suggested several options; however, in each case, the respondents emphasized the limitations inherent in their suggestions. A sample of their group options and comments is shown in **Exhibit 22**.

Nature of Comparison Group	Respondents' Quotes	Comments & Limitations
Random assignment of qualified training institute applicants to training (or not)	"We toyed with the idea of doing a random selection from our applicant pool to choose trainees. We don't know if this is unfair or unethical as a selection method. We thought about selecting the top 78-80 based on merit and then randomly selecting trainees from that top pool. We honestly don't know if it makes any	• This would be the strongest design in a prospective evaluation and would clearly answer the question of whether people who receive the training have better outcomes than those who do not.

Exhibit 22. NIH Program Officers' Suggested Comparison Groups

Nature of Comparison Group	Respondents' Quotes	Comments & Limitations
	difference to select on merit once you have narrowed it down to the top 80 or so."	 Random assignment is often considered to be fair and ethical when the number of treatment slots is limited. This would obviously not work for a retrospective evaluation.
Quasi-experiment Option #1: Comparison with unselected applicants	"We considered selecting a sample from the applicants who were not selected for the training. Clearly, there were people among the unselected applicants that were unqualified. It was obvious. So using them as a comparison group would be like comparing apples and oranges." "Is it fair to compare [training institute] trainees with those who applied and did not get selected? No." "Getting accepted versus rejected is an intervention in and of itself that could have an effect on people's careers."	 This would be the easiest type of comparison group to construct, but is subject to selection bias. It is highly likely that there would be a strong response bias for any attempt to conduct surveys or interviews with unselected applicants.
Quasi-experimental Option #2: Use propensity score matching to create a matched comparison from among unselected applicants	"One issue with propensity score matching is that we are dealing with a group of people that are not randomly selected from the population of university faculty. These were people who were motivated to submit an application. You could develop propensity scores based on academic achievement and background, but how would you measure the motivational or interest factor regarding [institute topic]? You would have to interview people . The psychological aspect would be missing from the predictive variables unless we had something with which to measure the psychological characteristics of the selected and unselected applicants prior to the training." "People [selection committee members] all have their own subjective way to select participants, even when criteria are specified. You get a huge variability, even when the criteria are quantified. No one has the time to evaluate 160 people at the level of detail that you would need. You end up dividing it up and it's not reliable. It can be garbage. What does it mean?"	 In addition to the specific problems noted by these comments, data on all selected and unselected applicants are not available for all years and for all training institutes. Propensity score matching involves the use of larger sample sizes than are available for these training institutes.

Nature of Comparison Group	Respondents' Quotes	Comments & Limitations
Quasi-experimental Option #3: Compare selected trainees from different training institutes	"Even if we wanted to compare [training institute] to other training institutes, it wouldn't work because [their trainees] are different from oursthere are also double and triple dippers [trainees who complete multiple training institutes]."	 The logic of this comparison is unclear. What would it mean if one training institute produced more favorable outcomes than other training institutes? One comparison could be among the three training institutes that addressed emerging topics, and among the four training institutes that addressed well- established topics. Differences within these groups could be explored in terms of how the institutes were taught. With the emerging training institutes, we found that for two of the three, specific Funding Opportunity Announcements had been issued. How would the presence or absence of this factor affect grant productivity?
		•
Quasi-experimental Option #4: Use of selected trainees as their own controls (pre-post analysis)	<i>"People can serve as their own comparison group and we can examine skills before the training and skills after the training."</i>	 This type of comparison would also be easy to construct. However, several training institutes targeted early stage investigators who might be at a disadvantage for this comparison. On the other hand, this could be very feasible for trainees at a middle or later career stage.
Quasi-experimental Option #5: Where there are FOAs for specific training institute fields, and compare individuals with awards to those who applied and did not receive awards.	"Look at people who applied for grants and got them versus those who applied and did not get them to see if participation in the training institute was a factor."	 This approach was used in one of the studies reviewed. Individuals who participated in a geriatric mental health training institute and had gone on to receive a K- award were compared with individuals who had received K-awards but did not participate in the training institute. However, only 60% of the training institute trainees had received K

Nature of Comparison Group	Respondents' Quotes	Comments & Limitations
		awards, thus making this a comparison between the most successful trainees and others.
Quasi-experimental Option #6 Identify individuals who did not apply for the training but who are at a similar point in their careers and from the same institution	"We could identify people who are at the same level of their career and similar to those who we are targeting, junior faculty." "Looking at those who attended versus those who did not that came from the same university to see if there are tangible differences regarding grant applications, career, etc."	 Difficult to identify people who did not apply who were at the same institution at the same time as the applicant. Doesn't control for interest and motivation,
Quasi-experimental Option #7: Compare different training institute cohorts	"The first year was almost a test group because it was so new. The Institute differed after that and the last one was our best one because we did refine as we went along." They could be compared against themselves like a repeated measures design." "I think the first year was such a fundamentally different year that it should be separated from the other two years. The two more recent years were more similar in many ways."	 The issue of historical bias is a potentially important consideration because individuals in earlier institutes have had longer time to attain outcomes, changes in how the institutes were structured and taught have occurred, changes in funding pay-lines at NIH have occurred (earlier years more difficult, later years easing up). The number of trainees within any single year cohort is relatively small (30-40 at most).

As this exhibit shows, a randomized trial would be the strongest option for a prospective evaluation design in the views of these respondents; however, it isn't an option here since the evaluation is to be retrospective in nature. The NIH Program Officers suggested seven different quasi-experimental options, but each one had serious limitations.

Responses from the seven training institute faculty members suggested similar comparison group options, including: prospective random assignment of selected trainees (2 respondents); unselected applicants (5 respondents); comparisons between training institutes (3 respondents); using trainees as their own controls (1 respondent); matched comparison groups based on successful and unsuccessful applicants for K awards (1 respondent); and comparison with matched peers at the same university and at similar point in careers (1 respondent). Of the five faculty members who mentioned unselected applicants, four indicated that this option was unsatisfactory. None of the faculty members we interviewed felt that any of the quasi-experimental options suggested was wholly satisfactory.

5.5.2 Pilot Study Using Trainees and Unselected Applicants

The evaluation team conducted a pilot study using data on trainees and unselected applicants from the 2011 cohort for the TIDIRH training institute. The purpose of this study was to test the feasibility of locating data on unselected applicants, and was limited to examination of the submitted and funded research grants and published research publications for trainees and unselected applicants during the two years following completion of their training. The samples consisted of 33 trainees who participated in the 2011 TIDIRH training institute, and a random sample of 33 trainees chosen from the unselected applicants. We found that trainees submitted and were awarded slightly more NIH grants than unselected applicants but the difference was not statistically significant. Trainees did publish significantly more articles than unselected applicants, and were more likely to be first authors on these publications. A major limitation of this analysis is that we were unable to determine which unselected applicants were wore similar to the trainees, and which ones were very unlikely to have been selected. We concur with the comments offered by our interviewees on the possible selection bias that may be present with unselected applicants.

5.6 Key Findings on Outcomes

This discussion of results on outcomes and the factors that drive them summarized results from our review of the evaluation literature on theoretical frameworks and past evaluations of NIH-sponsored summer research training institutes, and interview responses concerning the types of outcomes NIH Program Officers and SRTI faculty expected the seven training institutes to achieve. The objective in presenting this discussion was to identify the Summer Research Training Institute stakeholders' expectations for training institute outcomes and determine whether these expectations were similar to what would be suggested by training evaluation theory and past evaluations of NIH-sponsored summer research training institutes.

Key findings from this discussion include the following:

- The Kirkpatrick Four-Levels® Model is the dominant evaluation model used in the training evaluation field today. While it is not without its critics, the model is useful as a means of identifying the specific levels which evaluation of training programs should cover.
- An important feature of this model is the evaluation approach within which it is embedded. The Kirkpatrick® Model emphasizes the *contribution* of training to individual and organizational outcomes, rather than arguing that the effects of training are causal in nature. The evaluation task is to build a credible case that shows how the training contributes to various short-term, intermediate, and long-range outcomes. This case rests upon the creation of a chain of evidence, in which each link credibly supports the next. This chain includes both quantitative and qualitative data, and confirming each step in the chain leads to a convincing series of arguments demonstrating the value of the training.
- Based upon the types of outcomes identified by the SRTI NIH Program Officers and faculty, an outcome chain was constructed for the SRTIs as shown in Exhibit 23. This chain should be read from the bottom to the top, and shows the chain of evidence needed to demonstrate the value of the training.

• Qualitative data from trainee satisfaction forms can provide useful information on short-term (Kirkpatrick® Level 1 and 2) outcomes.



- Interview respondents were divided in their assessments of whether a comparison group was necessary for the evaluation of the SRTIs. Much of this uncertainty reflected considerable doubt concerning whether a quasiexperimental comparison condition was realistically feasible for the evaluation. While respondents suggested seven possible options, none were viewed as wholly satisfactory. Much of the uncertainty revolved around the issue of attribution versus contribution. While some respondents stated that they would like to be able to say that the training *caused* more favorable outcomes among those selected, there was recognition that this type of statement was not realistic unless there could be some form of random assignment of gualified applicants to training and non-training conditions. What could be argued, several respondents noted, was that the training *contributed* to more favorable outcomes for those who completed it. These respondents emphasized that the individuals who were selected to participate in the training institutes were highly qualified individuals who already showed strong indications that they would be likely to succeed in research careers. For that reason, it would make sense to focus on describing what the trainees ultimately achieved, and how they did it.
- There were, however, two options suggested by several respondents that might be feasible to include in such an analysis. The first was to compare training institutes within the two basic groups (those addressing better established scientific fields and methods, and those addressing the recently emerging fields). The value of this type of comparison is that it would allow for exploration of differences in outcomes based on variations in the structural elements of the training institutes. For the three training institutes that address emerging research fields, for example, two training institutes (ISSH and TIDIRH) also have FOAs that target new research on these topics, while the third training institute (mHealth) did not have a specific FOA until this year (2014). Examining research grant productivity for these three training institutes, could answer the question of whether providing the training by itself or in concert with specific FOAs stimulates new research grant productivity.
- The second option that might be feasible to include is a comparison between trainees' research grant and publication productivity prior to the training and afterward. This type of comparison would be especially useful for those trainees who were at middle or late career stages.

VI. Conclusions and Recommendations

This report has presented a wealth of data describing the background, goals, and histories of the seven trans-NIH Summer Research Training Institutes, their structure and the specific implementation processes they follow, and the types of outcomes that two major groups of stakeholders (NIH Program Officers and training institute faculty) expect from them. We have set these analyses against the backdrop of earlier evaluations of NIH-sponsored summer research training institutes. These studies have, for the most part, evaluated the training institutes from an accountability perspective—that is, did the training institutes meet their programmatic goals? This is certainly a minimal standard to meet, but a necessary threshold to cross in demonstrating the value of these training institutes to NIH.

The Feasibility Study addressed seven evaluation questions. Conclusions responsive to specific Feasibility Study questions are indicated by a text box that identifies the corresponding questions(s). Answers to the Feasibility Study questions are closely aligned to the recommendations for an outcome evaluation of the Summer Research Training Institutes, and are also included in the discussions that follow.

6.1 Feasibility of an Outcome Evaluation

The seven Summer Research Training Institutes have been operating for varying lengths of time ranging from 3-13 years. During this time, only one of the seven training institutes has received any type of evaluation. The total cost of funding these training institutes has been approximately \$15 million dollars, and increasing constraints at NIH on how funds can be used to support training activities has made it more difficult to obtain approvals to continue doing so. Tighter funding and the lengthy number of years some of the training institutes have been offered have led to discussions within senior NIH management about the continued need for and value of the SRTIs. In addition, there have also been discussions concerning whether the format of the SRTIs could be changed from their current residential format to an online approach, and whether it's reasonable for NIH to continue to bear the costs of providing these training institutes (versus enabling academic institutions to obtain funding to provide the training⁶).

These factors, and an interest in demonstrating what SRTIs have produced, led to OBSSR's current interest in conducting an outcome evaluation of the SRTI program. It is our conclusion, based on the evidence presented in this report, that an outcome evaluation of the SRTIs is both feasible and warranted, and we recommend that planning for an outcome evaluation should move forward.

Feasibility Study Question 1

Is an outcome evaluation of the seven trans-NIH Summer Research Training Institutes feasible and warranted at this time?

⁶ RFA-OD-13-009, *Short Courses on Innovative Methodologies in the Behavioral and Social Sciences* will provide funding in 2014 for several short educational programs offered by universities on topics overlapping the seven Summer Research Training Institutes.

6.2 Recommended Design for the Summer Research Training Institutes Outcome Evaluation

Based on consideration of the evidence presented in this report, we recommend that the outcome evaluation employ a mixed-methods multiple case study evaluation design. Details on the components of this design including a logic model, proposed comparison approaches, primary and secondary evaluation questions, and outcome measures and metrics are discussed in the following subsections.

6.2.1 Components of the Design

The Kirkpatrick Four Levels® evaluation model continues to dominate the training evaluation literature, and this model is proposed as the conceptual basis for the outcome evaluation. One limitation of the Kirkpatrick® Model in past evaluation studies has been the failure of many studies to adequately integrate the effects of the workplace

setting to which trainees return following training. The literature review found that most previous studies using the Kirkpatrick® Model have been situated in corporate or industrial settings. There are important differences between these settings and the academic research setting in which the majority of the training institute participants work. We obtained some sense of the types of what Kirkpatrick refers to as "required drivers" (supports, rewards, encouragement, reinforcement) from interviews with training institute faculty. By including them explicitly in the conceptual framework, we hope to provide a more refined operationalization of the Kirkpatrick® Model than is usually seen in published evaluations using the model. A logic model for the outcome evaluation is shown in **Exhibit 24.**

Feasibility Study Question 4

What is the optimal design for evaluating these training institutes?



Exhibit 24. Logic Model for the Trans-NIH Summer Research Training Institutes Outcome Evaluation

The design of the outcome evaluation is based on a multiple case study of the seven training institutes. Each training institute will be an individual case. Embedded within each of the seven cases will be three *Success Case Studies*. Both components are discussed below.

6.2.1.1 Multiple Case Studies

The seven case studies represent individual tests of the conceptual framework shown in the logic model. Our model posits a sequence of four levels of outcomes (as opposed to the usual three). The first level represents a merging of Kirkpatrick's® Levels 1 and 2 (reaction and learning), and includes outcomes such as increased self-confidence and motivation, increased understanding of the basic principles, theories and methods taught in each training institute, development of a personal plan for a research career, etc. These would be expected to represent the immediate results from participation in the training institute. The second level represents the period when the trainees return to their individual academic research institutions and begin to apply what they have learned in their own work, through pilot studies, new publications, grant applications, teaching, and new collaborations. These initial applications of what has been learned (Kirkpatrick's® Level 3) are likely to be affected by the types of supports and reinforcement offered by each trainee's institution as well as supports the trainees may obtain through maintaining contact with other trainees and faculty from the training institute.

As trainees apply what they learned (and obtain additional training), these results lead to the third level of the logic model, where increasing successes with publications, grants and teaching activities lead to career advancement, participation on NIH review panels, and movement into positions of increasing scientific activity and leadership in professional organizations. The accumulation of achievements at this third level in turn enable broader institutional outcomes in the forms of an expanded research portfolio at NIH and the growth of new training programs at universities. This fourth level represents the institutional impacts of the training institutes.

This model suggests a sequential relationship across the four levels that culminate in the overall Return on Expectation to NIH—the perception that the training institutes have created value in non-tangible but meaningful ways for NIH. We propose to test this model separately for each of the training institutes through the multiple case studies. To provide a fair test of these outcomes, we propose to establish quantitative metrics for each of the outcomes at each of the four levels of the model. We will work with the NIH Program Officers who have been involved in planning and developing each training institute to establish quantitative metrics for each of the outcome measures. These metrics will include a threshold target for each outcome. As an example, consider the outcome measure "evidence that a trainee is applying training material in his or her research activities (e.g., pilot studies, analyses of datasets, etc.)." A possible target threshold for this measure is that 60% of trainees will demonstrate this behavior within their first year following training. Measurement of this threshold can be performed by including appropriate items in a trainee survey. If data show that at least 60% of the trainees for a training institute have applied training materials in their research activities, the training institute will be considered to have obtained a "pass" for that outcome, and

will receive a score of 1 point on that measure. Each training institute will be scored in the same manner for all outcomes.

This strategy allows us to test the sequential nature of these outcomes—in order for a training institute to "pass" each level, it will be necessary to pass on the outcomes in the preceding level. It also allows us to apply the same yardstick to each of the seven training institutes.

6.2.1.2 Success Case Studies

Brinkerhoff's (2005) Success Case Method is a mixed methods evaluation approach that combines narrative case studies with quantitative methods through a fivestep process. To apply the method, it is first necessary to develop an impact model of the training, identifying the goals of the training and specifying what success looks like in light of these goals. Then, the evaluator conducts a survey of training participants to identify "best" and "worst" cases. The evaluator then selects a sample of best and worst trainees and obtains corroborating evidence from these trainees; this evidence usually derives from interviews but can include document review or other data that support the classification. The evidence focuses on how the trainees applied (or were unable to apply) the knowledge, skills, and attitudes learned in the training. The resulting data are analyzed to identify success factors, barriers, and challenges in transferring the learning into the worksite. Additional data can be obtained from other sources such as supervisors.

The purpose of the Success Case Studies is to examine in close detail how each training institute and the trainees' academic institution influenced the trainees' career development at each of three stages: early stage (within 10 years of the most advanced research degree), middle, and late stage. While the multiple case studies proposed above show what the level of accomplishment was for the various types of outcome measures, the Success Case Studies provide a detailed examination of how the different pieces (training institute outcomes and academic research setting supports) fit together and contribute to the trainees' careers at different stages.

6.2.2 Comparison Strategies Used in the Outcome Evaluation Design

The analysis of the NIH Program Officers and training institute faculty members' interview responses showed strong support for conducting a prospective randomized trial of the seven training institutes. To do this, each training institute's selection committee would need to identify a pool of qualified applicants for each

Feasibility Study Questions 5, 5a, & 5b

What types of comparison groups are feasible to address the study questions?

training institute and randomly assign them to either participate in the training institute or not participate. This would provide the clearest test of whether individuals who were interested in the scientific field for each training institute performed better than those who were equally qualified to attend but did not. However, this approach would not be suitable for a retrospective evaluation of the results from past sessions of the training institutes, which is what we were asked to address. Interview respondents identified seven possible quasi-experimental comparison strategies that might be applied with the seven training institutes. None of these options was strongly endorsed by any of the respondents, and in fact, all respondents were highly skeptical about each of them. Several respondents actively questioned whether a comparison group was even necessary, preferring instead an approach that would provide richer and more detailed information about the trainees' subsequent activities.

The evaluation team does, however, believe that two comparison strategies would be useful because they address two factors that could affect the outcomes each training institute has attained.

The first factor is the maturity of the scientific field each training institute addresses. As noted previously, four training institutes address more established scientific fields— ATI, C&AD, RCT, and SW. There has been historical support for research in these fields at NIH, and the shared aims of these four training institutes focused on building upon what the trainees already knew and understood about each of the four fields, rather than teaching something that was completely new. By contrast, three training institutes (ISSH, mHealth, and TIDIRH) each addressed fields that are newly connecting with health and biomedical research. Here the aim was to establish a community of researchers while substantially increasing and improving the NIH research portfolios for these fields. For these reasons, the evaluation team is recommending that some comparisons of training institute outcomes be compared within these two groups of training institutes, but not between them.

The second factor is the career stage of the trainees. While all seven training institutes ultimately included trainees from early and mid-late career stages, some training institutes explicitly targeted early stage investigators. It is also likely that some components of the training institutes might work more effectively for early stage investigators than for mid to later stage investigators (e.g., mentoring). This further supports the proposed use of the Success Case Studies approach since they provide a mechanism by which we can examine how the training experience and subsequent academic research institution supports and reinforcement work together to shape career trajectories.

We recommend against the other quasi-experimental comparison options discussed in the last section. These are not feasible either because the required assumptions are not met by the available data (propensity score matching) or because the inherent selection biases present in them would undercut the validity of the comparison.

6.2.3 Primary and Secondary Outcome Evaluation Questions

A list of the primary and secondary outcome evaluation questions proposed for the outcome evaluation is shown in **Exhibit 25**.

Feasibility Study Questions 2, 2a & 2b

What are the study questions an outcome evaluation should address?

Question Number	Primary and Secondary Evaluation Questions (Questions Are Tied To Specific Sections Of Logic Model, Exhibit 25)	Proposed Comparison (If Any)		
Training Institute Inputs				
1	What were the total training institute costs for the years included in this evaluation?			
2	What are the expected outcomes for the seven training institutes? (Note: This question involves establishing specific metrics/thresholds for each of the outcomes with the NIH Program Officers)			
	Training Institute Activities			
3	What were the characteristics of the trainees selected for each training institute?			
	Include the following variables and examine by year for each training institute:			
	 Career stage (early stage [within 10 years of advanced degree], middle stage [Associate Professor] and late stage [Professor]) NIH research grants during two years prior to training institute Publications during two years prior to training institute Trainee gender 			
4	How many trainees attended more than one training institute? Which training institutes?			
4A	Which institutions accounted for the largest numbers of trainees overall and by training institute?			
4B	Were institutions that accounted for the largest numbers of trainees more likely to establish their own training programs in those fields?			
4C	What was the Carnegie research classification for the home academic institutions to which trainees returned? Was Carnegie Research Classification associated with the number of trainees who attended training institutes?			
	Immediate Cognitive and Attitudinal Changes (Levels 1 and 2)			
5	What proportion of trainees from each training institute reported increased understanding of training content, increased ability to apply training content, increased motivation to pursue research in the training field, and other Level 1 and Level 2 outcomes? Did these outcomes vary by trainees' career stages?	Established versus Emerging Career Stage		
	Application of Training in Own Research Activities (Level 3)			
6	What proportion of trainees from each training institute reported applying training knowledge and skills to their research, teaching, and publication activities? Did these outcomes vary by trainees' career stages?	Established versus Emerging Career Stage		
7	How did attendance at a training institute affect subsequent scientific productivity (research grants and research publications)? Did trainees from ISSH and TIDIRH generate a greater number of research grants than those from mHealth?	Established versus Emerging Career Stage		
8	What proportion of trainees relocated to another academic institution following training?			

Exhibit 25. Primary and Secondary Outcome Evaluation Questions and Proposed Comparisons

Question Number	Primary and Secondary Evaluation Questions (Questions Are Tied To Specific Sections Of Logic Model, Exhibit 25)	Proposed Comparison (If Any)		
8A	What proportion of trainees who relocated reported that a major factor in the decision to relocate was that the new institution had a stronger academic program in the field taught by the training institute?			
8B	At what career stage were trainees who relocated more likely to do so?			
	Home Institution Required Drivers			
9	What types of research career supports and reinforcement did trainees' home academic institutions provide when they returned from the training?	Established versus Emerging Career Stage		
9A	How helpful did trainees perceive these research career supports and reinforcement?	Success Case Studies		
9B	Did research career supports vary by career stage?	Success Case Studies		
9C	What types of research career supports do trainees wish they had been provided at the time they returned from the training?	Success Case Studies		
9D	How did these research career supports and reinforcements contribute to career success?	Success Case Studies		
	Contextual Factors			
10	For which training institutes did NIH engage in additional activities to support development of research in this area, and how closely coordinated were these activities with the training institutes?	Established versus Emerging		
11	What were the NIH grant success rates over this period (FY 2001-FY 2014)?			
12	What other contextual factors might have affected trainees' success for each training institute?			
	Individual Career Impacts (Level 4)			
13	How did attendance at training institutes affect trainees' subsequent career advancement?	Established versus Emerging Career Stage		
Institutional (University) Impacts				
14	How many institutional training programs have been created in each field over the study period? How many universities developed new degrees or degree concentrations in the scientific fields over the study period?	Established versus Emerging		
NIH Impacts				
15	How many and what types of research and project grants have been awarded in each of these scientific fields since each training institute began?	Established versus Emerging		
16	How many new Funding Opportunity Announcements have been issued in each scientific field since the respective training institute began?	Established versus Emerging		

6.2.4 Outcome Measures, Metrics, and Data Sources

Specific outcome measures and metrics are listed in **Exhibit 26.** This Exhibit also shows the data source proposed for each measure. Specific data sources for the outcome evaluation will include NIH databases (IMPAC II, SPIRES, NIH RePORTER, internal files), data from past training

Feasibility Study Questions 3

3. What are the appropriate outcome measures and metrics for evaluating the training institutes?

institute satisfaction questionnaires, a proposed online survey of past trainees, and success case interviews with a sample of early and mid/late stage trainees.

Key Variables (Primary Or Secondary Evaluation Question Shown In Parentheses)	Metrics	Data Source
Cost per Training Institute per Year (1)	Total dollar costs for each training institute by Fiscal Year	NIH records
Expected Outcomes for the Training Institutes (2)	NIH Program Officers quantitative threshold for each outcome measure	Program Officer interviews
	(Example: within one year from the training institute, 25% of trainees will have submitted one or more NIH research grant applications)	
Trainee Characteristics: Career Stage (3)	Early Stage (within 10 years of year of highest research degree) Middle or Late Stage (more than 10 years after year of highest research degree)	NIH IMPAC II QVR
Trainee Characteristics: Research Grants During Past 24 Months (3)	Count of number (and type) of research grants awarded within the past 24 months before training institute	NIH IMPAC II QVR
Trainee Characteristics: Research Publications During Past 24 Months (3)	Count of number of research publications published during the past 24 months before training institute	NIH IMPAC II QVR
Trainee Characteristics: Gender (3)	Trainee gender	NIH IMPAC II QVR
Participation in Multiple Training Institutes (4)	Number of trainees attending more than one training institute	NIH records
Institutions (4A & 4B)	Institutions from which trainees were selected	NIH records
	Number of trainees participating in each training institute	
Carnegie Research Classification (4C)	Carnegie Research Classification of academic institution from which trainee came	Internet
Total Level 1 & 2 Pass Score (5)	Sum of the number of pass scores for Level 1 and 2 outcomes for each training institute	Trainee Online Survey

Exhibit 26. Key Variables, Metrics, and Data Sources

Key Variables (Primary Or Secondary Evaluation Question Shown In Parentheses)	Metrics	Data Source
Increased Understanding of Principles, Theory and Methods (5)	Response to Trainee Online Survey item	Trainee Online Survey
Increased Ability to Apply Principles, Theory and Methods (5)	Response to Trainee Online Survey item	Trainee Online Survey
Increased Awareness of Additional Learning Resources (5)	Response to Trainee Online Survey item	Trainee Online Survey
Identification of Specific Research Question (5)	Response to Trainee Online Survey item	Trainee Online Survey
Increased Motivation and Self-Efficacy (5)	Response to Trainee Online Survey item	Trainee Online Survey
Development of Personal Plan (5)	Response to Trainee Online Survey item	Trainee Online Survey
Total Level 3 Pass Score (6)	Sum of the number of pass scores for Level 3 outcomes for each training institute	Trainee Online Survey
Application of Training to Research Activities (6)	Response to Trainee Online Survey Item	Trainee Online Survey
Application of Training to Research Grant Applications (6)	Response to Trainee Online Survey Item	Trainee Online Survey
Application of Training to New Publications (6)	Response to Trainee Online Survey Item	Trainee Online Survey
Application of Training to Teaching Activities (6)	Response to Trainee Online Survey Item	Trainee Online Survey
Implementation of Personal Plan (6)	Response to Trainee Online Survey Item	Trainee Online Survey
New Research Collaborations (6)	Response to Trainee Online Survey Item	Trainee Online Survey
Contact with Other Trainees & Faculty (6)	Response to Trainee Online Survey Item	Trainee Online Survey
Use of Trainee Listserv, Post-training Webinars, Social Media (6)	Response to Trainee Online Survey Item	Trainee Online Survey
Subsequent NIH Research Grants (7)	Number of NIH Research Grants obtained after the training institute (by type and by year)	NIH IMPAC II QVR Trainee Online Survey
Subsequent non-NIH Research Grants (7)	Number of Research Grants from other sources obtained after the training institute (by type and year)	Trainee Online Survey
Subsequent Research Publications (7)	Number of Research Publications in training institute field by type and year)	SPIRES
Key Variables (Primary Or Secondary Evaluation Question Shown In Parentheses)	Metrics	Data Source
--	--	---
Subsequent Research Publications Field Percentage (7)	Percentage of total trainee publications that address training institute field	SPIRES
Relocation (8)	Number of trainees who obtained a new position at another university	Trainee Online Survey
Stronger Academic Program (8A)	Response to Trainee Online Survey Item	Trainee Online Survey
Research Career Supports (9)	Response to Trainee Online Survey Item	Trainee Online Survey Success Case Studies
Helpfulness of Research Career Supports (9)	Response to Trainee Online Survey Item	Trainee Online Survey Success Case Studies
Desired Research Career Supports (9)	Response to Trainee Online Survey Item	Trainee Online Survey Success Case Studies
Contribution of Research Career Supports (9)	Response to Trainee Online Survey Item	Trainee Online Survey Success Case Studies
Contextual Factors: Other NIH Activities (10)	Types of programmatic activities NIH pursued to promote research in scientific field	Program Officer Interviews
NIH Grant Success Rates (11)	NIH Success Rates by type of grant by Fiscal Year, 2002-2014	NIH IMPAC II QVR
Departmental or Institutional Awards (13)	Response to Trainee Online Survey Item	Trainee Online Survey
Contribution of Training Institute to Awards (13)	Response to Trainee Online Survey Item	Trainee Online Survey
Academic Promotions (13)	Response to Trainee Online Survey Item	Trainee Online Survey
Contribution of Training Institute to Promotions (13)	Response to Trainee Online Survey Item	Trainee Online Survey
Participation on NIH Grant Review Panels (13)	Response to Trainee Online Survey Item	Trainee Online Survey
Contribution of Training Institute to Participation on Review Panel (13)	Response to Trainee Online Survey Item	Trainee Online Survey
Activity in Professional Research Organizations (13)	Response to Trainee Online Survey Item	Trainee Online Survey
Scientific Leadership in Field (13)	Response to Trainee Online Survey Item	Trainee Online Survey
New Institutional Training Programs (14)	Response to Trainee Online Survey Item	Trainee Online Survey

Key Variables (Primary Or Secondary Evaluation Question Shown In Parentheses)	Metrics	Data Source
New Institutional Degree Programs or Concentrations (14)	Response to Trainee Online Survey Item	Trainee Online Survey
NIH Research Portfolio : Size (15)	Number of active NIH research grants in training field	NIH IMPAC II QVR
NIH Research Portfolio: Types of Projects (15)	Number of active NIH research grants in training field by type of grant	NIH IMPAC II QVR
New Funding Opportunity Announcements (16)	Number of Funding Opportunity Announcements issued for each training institute field	NIH IMPAC II QVR

6.2.5 Data Sources

The outcome evaluation will use data from four sources as shown in **Exhibit 27**. These sources are discussed below.

6.2.5.1 NIH Databases and Internal Records

The outcome evaluation will require data abstraction from NIH's IMPAC II QVR

database, which contains information on grant applications and awards and Principal Investigators. IMPAC II QVR is updated regularly and provides valid and reliable information on grant applicants, grant applications, and grant awards. It also contains information on research publications

Feasibility Study Questions 6 & 7

- 6. What are the available data sources, and what are their strengths and limitations?
- 7. What new data collection instruments will be needed?

associated with specific research grants. The outcome evaluation will also use the SPIRES database for research publications.

In addition to these NIH databases, the evaluation will also utilize internal records maintained on each training institute. Examples of data to be obtained from this source include cost data on each training institute, lists of trainees selected for each of the institutes during the years they were offered, and past trainee satisfaction data

Exhibit 27. Outcome Evaluation Data Sources

- NIH databases and internal records
- NIH Program Officer interviews
- Trainee online survey (new instrument)
- Success Case Study interviews (new instrument)

from each training institute. The pilot analysis of the ISSH satisfaction data showed that these data can provide useful insights into trainees' reactions. They can also furnish important information on the personal goals and objectives of trainees for their training experiences. The format and contents of these satisfaction data vary for different training institutes because each one used its own questionnaire. However, there are some common types of questions that can be examined across the institutes (especially whether trainees found the training to be relevant to their research needs.)

6.2.5.2 NIH Program Interviews

The primary purpose of the Program Officer interviews will be to obtain their input concerning the individual target thresholds for each of the outcome measures. This is a process that could easily be accomplished through a group meeting, especially since we want to obtain agreement on the part of the Program Officers concerning what constitutes a realistic target threshold for the various measures. Alternatively, this data collection could be done online as a modified Delphi process in which the individual Program Officers first establish thresholds for each item. The range, average, and median levels for the targets for each item would then be reported back to the participants in a second round at which time Program Officers could revise their initial targets as desired; the average of this second round would be taken as the target for each outcome.

6.2.5.3 Trainee Online Survey

A critical source of data for the outcome evaluation will be the trainees themselves. The trainee online survey will provide important information about trainees' reactions to training, what they learned, types of research supports their home academic institutions provided to them, their subsequent scientific activities and their career progress. One major challenge for using a survey in this design is that for some trainees, many years have passed since they participated in the training institute. By limiting the trainees surveyed to those within six years of training, we hope to minimize some of the possible problems with recall that trainees who attended many years ago would face. A second approach we will use is to place specific timeframes around some of the questions (for example, "During the first year after you completed the _____training institute, did you....."). This will offset a problem identified in our review of past summer research training institutes. In those studies that used a single survey for multiple training institute cohorts, the absence of a specific timeframe for survey questions meant that some individuals may have reported on several years while others reported for only a few months.

A draft copy of the Trainee Online Survey is attached (Appendix).

6.2.5.4 Success Case Study Interviews

The Success Case Studies (SCS) component will enable us to conduct intensive interviews with a small sample of trainees from each of the seven training institutes. The goal of the SCS is to explore how the trainees utilized the training they received and how the training in conjunction with the institutional research supports and research milieu in which the trainee operated contributed to a successful research career. The Success Case Studies will be conducted with investigators who took the training as early-stage investigators, or as middle-or late stage investigators to permit analysis of the different benefits that investigators in each of these two groups may have experienced.

Success Case Studies use a purposive sample based on the "…very best that the training is producing" in order to demonstrate clearly the contribution that training has had on the trainee. We will establish specific criteria for selecting this sample in consultation with the outcome evaluation's Evaluation Advisory Committee. As an

example, the criteria could include a requirement that the sampled trainees will be those who have obtained one or more NIH research grants, published one or more research articles within the field for which they received training, and who completed the training at least some specified number of years earlier.

The interview protocol used for the Success Case Study will inquire about what the trainee expected to achieve by taking the training, what supports the trainee home institution provided him or her on his or her return, how the trainee applied what was taught, what additional training (if any) he or she obtained in the field, teaching activities, subsequent scientific activities and career advancement. We would expect to see a pattern of outcomes very similar to those described by our logic model.

The Success Case Studies component and the broader multiple case studies work closely together to provide compelling data that demonstrate the value of the training institutes. The Success Case Studies provide these data at an individual trainee level, while the broader multiple case studies provide these data at a more aggregated level for each training institute. Together the two components provide a detailed picture of whether the training institutes are providing a return on NIH's expectations for them and how they do so.

References

- 1. Brinkerhoff RO (2005). The Success Case Study Method: A strategic evaluation approach to increasing the value and effect of training. *Advances in Developing Human Resources*, 7(1): 86-101.
- 2. Dunet DO and Reyes M (2006). Stakeholder-focused evaluation of an online course for health care providers. *Journal of Continuing Education in the Health Professions,* 26(4): 257-267
- 3. Ford JK, Kraiger K, and Merrett S (2010). An update review of the multidimensionality of training outcomes: New directions for training evaluation research. In Kozlowski SWJ and Salas E (Eds.) *Learning, Training, and Development in Organizations.* New York: Taylor-Francis.
- 4. Glasgow RE, Vinson CA, Chambers D, Khoury MJ, Kaplan RM, and Hunter C (2012). National Institutes of Health approaches to dissemination and implementation science: Current and future directions. *American Journal of Public Health*, 102(7): 1274-1281.
- 5. Holton EF III (1996). The flawed four-level evaluation model. *Human Resource Development Quarterly*, 7(1): 5-21.
- 6. Kaufman R, Keller J, and Watkins R (1995). What works and what doesn't: Evaluation beyond Kirkpatrick. *Performance & Improvement,* 35(2): 8-12.
- 7. Kirkpatrick DL (1993). *Evaluating Training Programs: The Four Levels.* San Francisco: Berrett-Kohler.
- 8. Kirkpatrick J and Kirkpatrick WJ (2009). The Kirkpatrick Four Levels [™]: A Fresh Look After 50 Years (1959-2009). Accessed at <u>http://www.managesmarter.com/managesmarter/images/pdfs/trg_20090417_kirk</u> <u>patrickw hitepaper.pdf</u>
- 9. Knowles M (1996). Adult learning. In Craig RL (Ed.) *The ASTD Training and Development Handbook.* New York: McGraw Hill.
- 10. Kraiger K, Ford JK, and Salas E (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology Monograph,* 78(2): 311-328.
- 11. Lemire S (2010) *Contribution analysis: The promising new approach to causal claims.* Paper presented at the European Evaluation Society Annual Conference, Prague, Czechoslavakia.
- 12. Maholmes V and Lomonaco CG (Eds.) *Applied Research in Child and Adolescent Development: A Practical Guide.* New York: Taylor and Francis, 2007.

- 13. Manning FJ, McGeary M, and Estabrook R (Eds). *NIH Extramural Center Programs: Criteria for Initiation and Evaluation.* Washington DC: National Academies Press, 2004.
- 14. Mayne J (2008) Contribution analysis: An approach to exploring cause and effect. *Institutional Learning and Change Brief No. 16.*
- 15. McLean S and Moss G (2003). They're happy, but did they make a difference? Applying Kirkpatrick's framework to the evaluation of a national leadership program. *Canadian Journal of Program Evaluation*, 18(1): 1-23.
- 16. McLinden D and Trochim WMK (1998). Getting to parallel: Assessing the return on expectations of training. *Performance & Improvement* 37(8): 21-26.
- Meissner HI, Glasgow RE, Vinson CA, Chambers D, Brownson RC, Green LW, Ammerman AS, Weiner BJ, and Mittman B (2013). The U.S. Training Institute for Dissemination and Implementation Research in Health. *Implementation Science*, 8: 12 doi:10.1186/1748-5908-8-12
- 18. National Institutes of Health (2003). *NIH Plan for Social Work Research*. Report prepared for the United States Congress.
- 19. Nilsen W, Kumar S, Shar A, Varoquiers C, Wiley T, Riley WT, Pavel M and Atienza AA (2012) Advancing the science of mHealth. *Journal of Health Communications*, 17(1): 5-10.
- 20. Nickols FW (2005). Why a stakeholder approach to evaluating training. *Advances in Developing Human Resources,* 7(1): 121-134.
- 21. Phillips JJ (1997). *Return on Investment in Training and Performance Improvement Programs.* Houston: Gulf Publishing.
- 22. US Office of Personnel Management (2011). *Training Evaluation Field Manual— Demonstrating the Value of Training at Every Level.* Washington DC
- 23. Wang GG and Wilcox D (2006). Training evaluation: Knowing more than is practiced. *Advances in Developing Human Resources*, 8: 528-539.

Appendix

Proposed Online Trainee Survey Instrument

NIH Summer Research Training Institutes—Online Trainee Survey

TRAINEE CHARACTERISTICS

1. Which of the following best describes your **primary** organizational affiliation? (Please select one.)

Academic department within a college or university

1a. IF YES: Are you also affiliated with a research center within this university? ☐ Yes

 \square No

Research position within a non-profit organization

Research position within a for-profit organization

Medical or clinical services

Other (Please specify....)

2. Which of the following best describes your current position? (Please select one.)

Faculty

Research Staff Scientist

Postdoctoral Fellow

Other (Please specify....)

- 3. What is your primary department or field? (Please select one.) [DROP DOWN BOX]
- 4. What research degrees do you hold? (Please check all that apply.)

	H
	ľ
	6

PhD (Please specify year....) MD (Please specify year....) Other (Please specify degree(s) and year(s)

5. What is your gender?

Male Female

NIH SUMMER RESEARCH TRAINING INSTITUTE EXPERIENCE

6. Which NIH Summer Research Training Institute(s) did you attend? (Please check all that apply.)

NIH SUMMER RESEARCH TRAINING INSTITUTE	. YEAR TAKEN
Advanced Training Institute on Health Behavior Theory	
Institute on Systems Science and Health	
mHealth Training Institute	
NIH Summer Institute on Social and Behavioral Intervention Research	
Randomized Clinical Trials involving Behavioral Interventions	
Summer Institute in Applied Research in Child and Adolescent Development	
Summer Training Institute on Dissemination and Implementation Research in Health	

CONSIDER THE MOST RECENT TRAINING INSTITUTE YOU ATTENDED IN ANSWERING THE FOLLOWING QUESTIONS. (This can be programmed into the survey so that the most recent training institute identified in the above question appears in the blanks for the following questions.)

7. How did you first learn about the _____ NIH Summer Research Training Institutes? (Please select all that apply.)

NIH wel	bsite
---------	-------

NIH Program Officer

- Colleague who had attended this training institute in a previous year
- Scientific meeting
- Faculty mentor
- Other (Please specify [DROP DOWN TEXT BOX])
- 8. Why did you attend the ______NIH Summer Training Research Institute(s)? (Please check all that apply.)

I was interested in the topic area

Felt that I needed to go to enhance my career opportunities

Interested in expanding into a new research area

Interested in meeting new people in that particular field

Other (Please specify....)

9. Looking back on your experience with the _____Summer Research Training Institute from your perspective today, how would you rate your overall satisfaction with the following aspects of the experience?

	Poor	Below	Average	Above	Exceptional	N/A
		Average		Average		
Topics covered (course content)	1	2	3	4	5	NA
Lectures	1	2	3	4	5	NA
Opportunity to work on your own project	1	2	3	4	5	NA
or idea						
Interactions with faculty	1	2	3	4	5	NA
Interactions with other trainees	1	2	3	4	5	NA
Opportunity to work with a mentor	1	2	3	4	5	NA
Opportunity to immerse yourself in the	1	2	3	4	5	NA
training						
due to the residential element						
Other (Please specify)	1	2	3	4	5	NA

10. How much did the NIH Summer Research Training Institute contribute to each of the following? (Please fill in one response for each row.)

	Very Little				Very Much
Gaining new knowledge	1	2	3	4	5
Gaining new skills	1	2	3	4	5
Making new contacts	1	2	3	4	5
Developing new relationships	1	2	3	4	5
Improving my ability to write a grant	1	2	3	4	5
Knowing who to talk to at NIH if I need	1	2	3	4	5
grant assistance					
Improving my confidence to write a grant	1	2	3	4	5
Improving the likelihood that I would submit a grant	1	2	3	4	5
Increasing my awareness of where to go to learn					
more about the training institute topic					
Other (Please specify)	1	2	3	4	5

- 11. At the time you participated in the _____Summer Research Training Institute, did your academic department, research center, or organization promise to provide you with any type of assistance or support to further your research career?
 - □ Yes (Complete Questions 11a and 11b)

 \Box No (SKIP to 12)

11a. At the time you participated in the _____ Summer Research Training Institute, which of the following types of post-training support did your academic department, research center or organization promise to provide to you to help you apply what you had learned in the training?

τνρε ος εμβρορτ	My Department or Organization:		
I IFE OF SUFFORT	Provided	Did not	
	this	provide this	
Reduction in teaching or clinical workload to allow time for research			
Funding for pilot study or preliminary data collection to support grant			
development			
Access to a research assistant			
Access to laboratory space or equipment I did not have access to before			
Funds to travel to present research at a conference or professional meeting			
Opportunities to share new knowledge or skills within my department or			
center			
Purchased new equipment or software that would support my research			
Access to administrative assistance for preparation of a grant			
Access to statistical consultation			
Encouraged to work on grant applications and research			
Opportunities to include the knowledge and skills I had gained in my			
teaching activities			
Other (please describe)			

11b. Upon your return from the _____Summer Research Training Institute, what was the single most important type of support or assistance that you received from your academic department, center, or organization?



12. In which of the following areas have you applied the knowledge and skills you learned at the _____Summer Research Training Institute? (Check all that apply)

(Note: Survey will be programmed such that respondent only views questions for the categories checked).

- □ Research
- □ Publications
- □ Teaching
- □ Professional Activities
- □ Other (Please explain)

 $\hfill\square$ None of the above

FIRST YEAR APPLICATION OF SUMMER RESEARCH TRAINING INSTITUTE KNOWLEDGE AND SKILLS

The following question asks you to think about what you did to apply what you had learned at the ______Summer Research Training Institute **during the first year after you returned to your academic department, research center, or organization**.

13. **During the first year after you had completed** the ____Summer Research Training Institute, which of the following actions did you take (check all that apply)?

First Year Activities	YES	NO
Developed a new idea for research based on what I learned from the training		
institute.		
Submitted a new research grant application based on what I had learned at the		
training institute.		
Contacted a faculty member I had met during the training institute to discuss		
issues related to the training content.		
Contacted other trainees I had met during the training institute to discuss ideas		
related to the training content.		
Contacted a NIH Program Officer to discuss ideas about a potential research grant		
application based on what I learned at the training institute.		
Collaborated with other trainees I had met during the training institute on one or		
more research grant applications.		
Collaborated with other trainees I had met during the training institute on one or		
more research publications.		
Relocated to another university that had a stronger program in the training		
institute content area.		
Implemented a plan for my research career that I had developed during the		
training institute.		
Gave a seminar on an aspect of what I had learned at the training institute.		
Incorporated content from the training institute into a course or courses that I		
teach.		
Conducted a pilot study based on what I learned at the training institute in order		
to collect data for a research grant application.		
Joined an existing multidisciplinary research team conducting research on a topic		
related to what I learned at the training institute.		
Formed a new multidisciplinary research team to conduct research on a topic		
related to what I learned at the training institute.		
Introduced content from the training institute to individuals I mentor.		

14. Please rate the degree to which you felt that your academic department, research center, or work unit within your organization encouraged and supported your efforts to apply the knowledge and skills you learned in the ____Summer Research Training Institute:

Very Little		Very Much

The following questions apply for **all of the years since you completed the ____Summer** Research Training Institute

RESEARCH ACTIVITIES RELATED TO ______SUMMER RESEARCH TRAINING INSTITUTE

- 15. Have you *submitted* any research grant applications related to the topic of the ____Summer Research Training Institute to one or more of the following agencies or organizations? (Check all that apply)
 - □ National Institutes of Health
 - □ Agency for Healthcare Research & Quality
 - □ Department of Veterans Affairs
 - □ National Science Foundation
 - □ Any other Federal agencies (Department of Defense, Department of Agriculture, etc.)
 - □ Any non-Federal foundations or organizations
 - □ I have not submitted any research grant applications related to this topic.
- 16. Have you been *funded* for one or more research grants related to the topic of the _____Summer Research Training Institute by any of the following agencies or organizations? (Check all that apply)
 - □ National Institutes of Health
 - □ Agency for Healthcare Research & Quality
 - □ Department of Veterans Affairs
 - □ National Science Foundation
 - □ Any other Federal agencies (Department of Defense, Department of Agriculture, etc.)
 - □ Any non-Federal foundations or organizations
 - □ I have not received funding for any research grants related to this topic.

- 17. Have you *participated on* any research grants led by someone else that related to the topic of the _____Summer Research Training Institute funded by one or more of the following agencies or organizations? (Check all that apply)
 - □ National Institutes of Health
 - □ Agency for Healthcare Research & Quality
 - □ Department of Veterans Affairs
 - □ National Science Foundation
 - □ Any other Federal agencies (Department of Defense, Department of Agriculture, etc.)
 - □ Any non-Federal foundations or organizations
 - □ I have not participated on any research grants related to this topic.
- 18. Did you collaborate with any of the trainees or faculty you met during the ____Summer Research Training Institute on one or more:
 - □ Research grant applications you submitted that were not funded?
 - □ Research grant applications someone else submitted that were not funded?
 - □ Research grants you submitted that were funded?
 - □ Research grants someone else submitted that were funded?

PUBLICATION ACTIVITIES RELATED TO _____SUMMER RESEARCH TRAINING INSTITUTE

19. Have you made any oral presentations before professional or scientific societies or organizations on topics related to what you learned at the _____Summer Research Training Institute?

	Yes (IF	YES: How	many pres	entations?)
--	---------	----------	-----------	-------------

🗌 No

19a. IF YES: Did you collaborate with any of the contacts you made at the Summer Research Training Institute in the preparation of the presentation(s)?

🗌 Yes

No No

20. Have you published any peer-reviewed research articles based on topics related to what you learned at the _____Summer Research Training Institute?

Yes (IF YES: How many articles?)
🗌 No
20a. IF YES: Did you collaborate with any of the contacts you made at the Summer Research Training Institute in the preparation of the publication(s)?
Yes
🗌 No
Have you published any books, book chapters, monographs, or other types of professional non-peer-reviewed publications based on topics related to what you learned at theSummer Research Training Institute?
Yes (IF YES: How many publications?)
🗌 No
21a. IF YES: Did you collaborate with any of the contacts you made at the Summer Research Training Institute in the preparation of the publication(s)?
Yes

🗌 No

21.

22. Do you engage in any teaching activities as part of your professional responsibilities?

Yes

No No

22a. Which of the following statements describes how you have integrated material from the _____Summer Research Training Institute into your teaching activities?

	YES	NO
I have integrated materials and topics from the training institute into		
courses I have taught or am teaching.		
I have used materials from the training institute in my mentoring		
activities with junior faculty or postdoctoral fellows.		
I teach one or more courses that are part of a specific academic training		
program or course concentration that focuses on this training institute		
topic.		
I developed an academic training program or concentration that		
focuses on this training institute topic.		

- 23. Does your university offer an academic program or course concentration that focuses on this training institute topic?
 - Yes (Complete 23a)
 - \square No
 - Did your participation in the _____ Summer Research Training Institute contribute to 23a. the development of this training program or course concentration?
 - □ My participation in the ____Summer Research Training Institute contributed *substantially* to the development of this training program or course concentration.
 - □ My participation in the Summer Research Training Institute contributed *somewhat* to the development of this training program or course concentration.
 - □ My participation in the _____Summer Research Training Institute had *little or nothing to do* with the development of this training program or course concentration.

24. During the year(s) since you completed the _____Summer Research Training Institute, did you (select appropriate statement):

	YES	NO
Complete additional training in the topic addressed by the training		
institute?		
Seek but not find additional training in the topic addressed by the		
training institute?		
Decide that you had received sufficient training on the topic addressed		
by the training institute?		

- 25. Did your participation in the _____Summer Research Training Institute contribute to your obtaining tenure?
 - \Box Yes (Please explain how)
 - □ No, I already had tenure at the time I participated in the _____Summer Research Training Institute.
 - □ No, it did not contribute to my getting tenure.
- 26. Has participating in the _____Summer Research Training Institute contributed to advancing your research career beyond the tenure stage?
 - □ Yes (Please explain...)
 - □ No

PROFESSIONAL ACTIVITIES RELATED TO _____SUMMER RESEARCH TRAINING INSTITUTE

27. Since completing the _____Summer Research Training Institute, which of the following activities have you pursued (check all that apply):

	YES	NO
I have participated in one or more NIH Internal Review Group panels		
(either Center for Scientific Review or Special Emphasis Panels).		
I have served as a member of a Data Safety Monitoring Board for a		
clinical trial.		
I have joined a professional association or organization that is closely		
related to the topic of theSummer Research Training Institute.		
I have actively participated in organizing and running activities for a		
professional association or organization that is closely related to the		
topic of theSummer Research Training Institute.		

I have served as a manuscript reviewer for a journal that is closely
related to the topic of theSummer Research Training Institute.
I have served on the Editorial Board for a journal that is closely related
to the topic of theSummer Research Training Institute.
I have received awards or honors from my university or organization for
research I have conducted that is closely related to the topic of the
Summer Research Training Institute.
I have received awards or honors from a professional association or
organization for research I have conducted that is closely related to the
topic of theSummer Research Training Institute.

FINAL REFLECTIONS ON PARTICIPATING IN THE ____SUMMER RESEARCH TRAINING INSTITUTE

28. What do you believe was the single most important benefit that participating in the ______Summer Research Training Institute had for you personally/professionally?