Feasibility Study of Optimal Approaches for Evaluating the NIAID Good Clinical Practice Computer-Based Training (GCP CBT) Program

Prepared for:

Laura McNay, Contract Project Officer
Judy Zuckerman, DCR Project Manager
Karin Lohman, Acting Chief, SPEB
Nancy Jones, SPEB Project Manager
Melba Reed, Evaluation Specialist
Terrie Nester, Contracting Officer

Prepared by:

Paul Young, M.P.H., M.B.A.
Lisbeth Jarama, Ph.D.
Sue Ruckel, LCSW-C
Allison Zambon, M.H.S.

NOVA Research Company
4600 East-West Highway, Suite 700
Bethesda, MD 20814-6900

Contract Number: 263-FF-611593

May 31, 2007

Table of Contents

Executive Summary iii

Introduction 1

Purpose of Feasibility Study 2

Evaluation Plan Development 2

Overview of Evaluation Approach 2

Feasibility Study Questions 5

Theoretical Evaluation Framework 7

Evaluation Plan 7

Step 1 - Engage Stakeholders 7

Step 2 - Describe NIAID GCP CBT Program 8

2.1. Evaluation Program Theory (Logic Model) *8*

2.2. Outcomes of Interest to the NIAID GCP CBT Program *9*

Step 3 - Focus the Evaluation Plan 9

3.1. Purpose *9*

3.2. Users *9*

3.3. Uses *9*

3.4. Advisory Committee *10*

3.5. Evaluation Questions *10*

3.6. Benchmarks and Performance Indicators of GCP CBT for Evaluation Component A: Evaluation of GCP CBT Effects *11*

3.7. Timeline *12*

3.8. Evaluation Planning Matrix *12*

3.9. Methods *13*

Step 4 - Gather Credible Evidence 14

4.1 Data Sources *14*

4.2. Data Analysis *16*

Step 5 - Justify Conclusions 17

Step 6 - Ensure Use and Sharing of Lessons Learned 17

Answering Evaluation Questions on Component A: Evaluation of GCP CBT Effects 18

Key Variables 18

1.1. Program Resources *18*

1.2. Population Characteristics *18*

1.3. Program Activities *18*

1.4. External Factors *18*

1.5. Program Goals, Performance Measures, and Comparison Measures *19*
Answering Evaluation Questions on Component B: Examination of GCP Compliance 27
Key Variables 27
Program Goals, Performance Measures, and Comparison Measures 27
Summary 29
References 30
APPENDIX A: NIAID GCP CBT Logic Model 31
APPENDIX B: NIAID GCP CBT Evaluation Matrix 32
APPENDIX C: Proposed Evaluation Timeline 35
APPENDIX D: NIAID User Profile Questionnaire 39
APPENDIX E: GCP CBT User Satisfaction Survey 41
APPENDIX F: Summary of Recommendations for a Comprehensive Evaluation of the GCP CBT Program that Includes Evaluation Component A: *Evaluation of GCP CBT Effects* and Evaluation Component B: *Examination of GCP Compliance* 42
APPENDIX G: NIAID GCP CBT Review of Literature Summary 45
APPENDIX H: Summary of Interviews with Site Monitors 53

NIAID Good Clinical Practice
Computer-Based Training Program

Report on the Evaluation Feasibility Study

Executive Summary

The focus of this feasibility study is to identify appropriate evaluation methodologies, techniques, and tools to measure relevance, effectiveness, and impact of the National Institute of Allergy and Infectious Diseases (NIAID), Division of Clinical Research (DCR), Good Clinical Practice (GCP) Computer-Based Training (CBT) Program. As part of NIAID's continuing desire to conduct high-quality, efficient clinical trial research, and in keeping with National Institutes of Health (NIH) mandates on Human Subjects Protection (HSP) [Code of Federal Regulations, Title 45, Part 46] and Food and Drug Administration (FDA) mandates on Good Clinical Practice in FDA-regulated clinical trials [Code of Federal Regulations, Title 21], a program to conduct a systematic training in Good Clinical Practice including HSP was developed. This program, "The Good Clinical Practice Computer-Based Training Program" was developed by EduNeering in collaboration with NIAID staff and planned for implementation in the fall of 2007.

The overarching goal of NIAID/DCR's GCP CBT is to increase knowledge and use of GCP; that is, best practices in planning, initiation, conduct, closeout, and reporting of clinical trials involving human subjects. GCP CBT objectives include demonstrating value added (i.e., improvement) in three areas: (1) subject safety; (2) data integrity—study data that are more reliable, trusted, accurate, and complete; and (3) efficiency—bringing a research project from research questions to answers expeditiously.

Donald Kirkpatrick's (1975) model for evaluating training programs was the framework used to assess GCP CBT effects. According to this model, four levels of training can be evaluated: ***Reaction*** to training (thoughts and feelings about the training), ***Learning*** of content (increase in knowledge, capacity, skills), ***Transfer*** of learned content to trainee's setting (behavior change, use of new skills), and ***Results/Impact*** of transferred content in trainee's setting (effect on the setting/environment resulting from trainee's performance).

After conducting the Feasibility Study, NOVA has come to the recommended conclusion that a full-scale evaluation of the GCP CBT is possible and should be conducted to determine the effectiveness of the training in increasing knowledge of GCP among NIAID clinical researchers and staff. The feasibility study identified two main evaluation components associated with Kirkpatrick's model:

- **Evaluation Component A:** Evaluation of GCP CBT effects (*Reaction, Learning)*
  To assess the influence of CBT on knowledge acquisition of GCP.
- **Evaluation Component B**: Examination of GCP compliance *(Transfer, Impact)**
  To explore GCP compliance in clinical trials research over time.

* Note of caution*:* While behavior change in using GCP can be measured over time, it will not be possible to directly associate the GCP CBT with any measured change in compliance. However, it is still important to evaluate whether NIAID's increased emphasis on GCP from multiple activities is resulting in change in use and associated improvements in clinical research.

The evaluation components complement each other by providing a general picture of NIAID's focus on GCP compliance in the conduct of clinical trial research. Exhibit 1 displays key information that will be learned from both evaluation components.

EXHIBIT 1. KEY INFORMATION TO BE LEARNED FROM THE FULL-SCALE EVALUATION

From **Evaluation Component A**: Evaluation of GCP CBT Effects	From **Evaluation Component B**: Examination of GCP Compliance
✓ Level of awareness of GCP CBT among target users ✓ Current knowledge of GCP among users (e.g., strengths and weaknesses) ✓ Overall reaction to GCP CBT (e.g., satisfaction) ✓ Perceived relevance of GCP CBT to users' work setting ✓ Effectiveness of GCP CBT in increasing knowledge of GCP among users ✓ Suggestions to improve effectiveness of GCP CBT	✓ GCP compliance in NIAID-funded clinical trial research studies over time ✓ Barriers, facilitators, and lessons learned in adherence to GCP guidelines ✓ Improvement in subject safety, quality of research data, and overall research efficiency over time, after availability of GCP CBT ✓ Suggestions to facilitate GCP compliance (identified by type of clinical trial study—e.g., R34s, and study stage—e.g., initiation)

The information obtained through activities conducted as part of the feasibility study was used to support evaluation design strategies and methodologies recommended in this report. A summary of key recommendations for a full-scale evaluation is displayed in Exhibit 2.

EXHIBIT 2. KEY RECOMMENDATIONS FOR EVALUATION STEPS

Step 1- Engage appropriate stakeholders in the evaluation. Key stakeholders include NIAID leadership and staff within DCR and SPEB, members of the Evaluation Advisory Committee, other NIAID Divisions conducting clinical trials, and the clinical trial research community.

Step 2- Describe the program. Review the program logic model annually to reflect program changes.

Step 3 – Focus the evaluation:

- Form an Evaluation Advisory Committee with members with experience in evaluation (i.e., evaluation of GCP, HSP, e-learning, adult learning strategies), GCP, NIH clinical trial regulations, and monitoring.
- Identify Evaluation Questions for Component A: Evaluation of GCP CBT Effects. Questions address program awareness and usage, overall reaction to training, and learning of GCP.
- Identify Evaluation Questions for Component B: Examination of GCP Compliance. Questions explore adherence to GCP over time, related barriers and facilitators, and improvement in subject safety, quality of data, and overall research efficiency.
- Establish realistic benchmarks for GCP CBT process and outcome indicators.
- Allow enough time for a full-scale evaluation. A 5-year plan is an appropriate time period to assess both evaluation components (e.g., focus on intramural research the first 2 years and then assess extramural research; examine domestic clinical trials first and then international studies; and collect longitudinal data over a reasonable period to examine GCP use and improvements in the conduct of clinical trials).
- Use a mixed-method evaluation design of qualitative and quantitative techniques to obtain a full picture of GCP CBT

effects and GCP compliance.
Step 4 – Gather credible evidence: ➢ Develop and test data collection instruments with input of the Evaluation Advisory Committee and pertinent audiences (e.g., site monitors, program staff). ➢ For Evaluation A: Evaluation of GCP CBT Effects, collect data on participants' characteristics (e.g., role at NIAID, previous GCP experience), participation data (e.g., courses most frequently accessed, started, and completed by users), pre- and post-tests of knowledge, and satisfaction with training. These data can be collected through EduNeering's database, the GCP CBT User Satisfaction Survey, and focus groups. ➢ For Evaluation B: Examination of GCP Compliance, collect data on key GCP data elements, perceptions regarding barriers and facilitators of GCP use, subject safety, data quality, and overall research efficiency. These data can be collected through a Site Monitoring Tool, the Survey on Issues with GCP Compliance, focus groups, and interviews. Collect related data from site monitors, study staff, and NIAID representatives, where appropriate.
Evaluation Component A: Evaluating GCP CBT Program Goals
Goal 1: Target audiences are aware of GCP CBT ➢ The program dissemination plan should be very specific, describing the number of expected activities directed to whom, by whom, and by when, and identifying benchmarks for activities (e.g., minimum number of activities by type planned to implement).
Goal 2: GCP CBT users have favorable reactions to the training ➢ Conduct a pilot of GCP CBT including administration of evaluation instruments and focus groups. ➢ Set benchmarks for expected ratings on overall satisfaction with CBT. ➢ Collect data that will be of use to NIAID program staff and stakeholders.
Goal 3: GCP CBT increases knowledge of GCP ➢ Select an evaluation design that minimizes threats to validity of evaluation findings (cons). ➢ Consider an evaluation design that includes a comparison group with pre- and post-tests, as this can best attribute differences in GCP knowledge to CBT. ➢ Determine the expected effect size on knowledge change (realistic) to establish an appropriate sample size.
Evaluation Component B: Examination of GCP compliance
Goal 4: Increased application of GCP in conduct of clinical trial research Goal 5: Improved subject safety Goal 6: Improved quality and integrity of clinical research data Goal 7: Improved research efficiency ➢ Collect data on GCP compliance at least 3 times (e.g., baseline Year 1, end of Year 3, and end of Year 4) to better assess any change in GCP use. ➢ Collect baseline data before GCP CBT is implemented. ➢ Collect GCP compliance data on new clinical trial studies only, to better assess any potential influence of the program on GCP use. ➢ Focus on a specific type of clinical trial to examine GCP compliance.

This report summarizes NOVA Research Company's (NOVA's) approach to developing a GCP CBT evaluation plan, as well as theoretical foundations on which the evaluation approach was built. The report also provides tools created by NOVA—such as the GCP CBT Program Evaluation Logic Model and Evaluation Planning Matrix—that NIAID can use to develop a comprehensive evaluation plan. Exhibit 3 on the next page presents a compliance matrix summarizing how this report addresses each of the technical requirements of this project.

EXHIBIT 3. MATRIX OF FINAL REPORT COMPLIANCE WITH TECHNICAL REQUIREMENTS OF THE FEASIBILITY STUDY OF OPTIMAL APPROACHES FOR EVALUATING THE NIAID GCP CBT PROGRAM

Detailed Technical Requirement	Description	Report Section Where Addressed
Identify relevant stakeholders. Clarify issues and objectives. Develop a logic model.	Document those who have an interest in evaluation findings and the extent of their involvement in evaluation planning. Identify key stakeholders, GCP CBT goals and objectives, and related issues of relevance to evaluation of the program. Develop a logic model of the GCP CBT to facilitate shared understanding of the program's structure, resources, planned activities, and outcomes.	Step 1 and Step 2, pages 7 to 9; Appendix A: GCP CBT Logic Model.
Conduct a literature review. Develop study questions for the evaluation. Identify key variables.	Review relevant literature, related studies on GCP, and evaluations of human subjects research training. Identify evaluation questions on program process and outcomes of interest to stakeholders which are clear, specific, and answerable. Discuss specific information needed to answer study evaluation questions.	Step 2 and Step 3, pages 8 to 11; Answering Evaluation Questions, pages 17 to 28; Appendix B: Evaluation Matrix; Appendix G: Results of Literature Review.
Review existing data. Plan for data collection and analysis.	Review existing data sources to identify key variables for the evaluation. Determine types of data that will be used to answer study questions. Identify feasible performance and comparison groups. Develop necessary data collection instruments. Develop a plan for data analysis. Determine strategies to ensure data integrity and address ethical considerations.	Step 3, pages 12 to 16; Answering Evaluation Questions, pages 17 to 28; Appendix B: Evaluation Matrix; Appendix H: Interviews with Site Monitors
Recommend an evaluation design.	Submit an outcome evaluation design. Recommend for or against proceeding with process and outcome evaluation and provide justification.	Answering Evaluation Questions, pages 17 to 28; Summary, page 28.

NIAID Good Clinical Practice
Computer-Based Training Program

Report on the Evaluation Feasibility Study

Introduction

The National Institute of Allergy and Infectious Diseases (NIAID), Division of Clinical Research (DCR) is implementing a Good Clinical Practice Computer-Based Training (GCP CBT) Program. GCP is an internationally accepted standard for conducting clinical trial research. The GCP CBT was developed to provide comprehensive and standardized training for NIAID intramural and extramural research personnel on scientific and ethical standards of human subjects research including those of the National Institutes of Health (NIH), Food and Drug Administration (FDA), Health and Human Services (HHS), and international clinical trial policies, guidelines, and regulations.

Historically, training in GCP has been delivered through a variety of modalities including Barnett's 1-day seminars, on-the-job training, informal GCP instruction for NIAID site monitors and Program Officials, certification programs offered by organizations such as the Association of Clinical Research Professionals (ACRP) and the Society of Clinical Research Associates (SoCRA), and computer-based training programs such as the Collaborative IRB Training Initiative (CITI) program. There has not been a standardized approach to the training of GCP with approved, accepted, and standardized course content, particularly as the content relates to National Institutes of Health federally funded clinical trial policies, regulations, guidelines, and reporting requirements. Furthermore, there has been no mechanism to evaluate whether learning has occurred and whether learning has resulted in behavior change that leads to better research practices, conduct, and protection of human subjects. In accordance with the NIH Clinical Research Policy Analysis and Coordination Program (CRpac), NIH policy recommends that all key personnel involved in the design or conduct of clinical research receive education on GCP and Human Subject Protection (HSP) before NIH funds are awarded.

In the winter of 2005, NIAID leadership concluded that it would be beneficial to provide a centralized and efficient delivery method of GCP training for all NIAID staff, contractors, and grantees who are involved in the conduct of human subject research. A GCP CBT Task Force was assembled to select a contractor to implement the CBT training. The Task Force chose EduNeering, a solutions-based company with e-learning experience and compliance-based learning solutions for regulated industries, to design and implement the computer-based training. The training curriculum covers topics including clinical trial monitoring, institutional review board (IRB) reports, research protocols, and Investigational New Drug (IND) forms. These "off-the-shelf" courses were chosen to satisfy requirements of GCP training and the course curriculum was later modified to meet more specific needs of NIAID.

The NIAID GCP CBT initiative is also in line with NIH CRpac that states that *efficiency and effectiveness in the system of clinical research is hampered by variability in regulations and policies that pertain to the conduct and oversight of clinical research*. The standardization of GCP training aligns NIAID more closely with CRpac goals of training for more efficient and effective conduct of clinical trials.

NOVA worked collaboratively with representatives from NIAID's GCP CBT Workgroup, DCR, and Strategic Planning Evaluation Branch (SPEB) to develop the best overall approach and most appropriate measures to evaluate the GCP CBT. The statement of work for the feasibility study specified the overall approach for the evaluation:

> *"The GCP CBT is a new NIAID program and both DCR and SPEB are interested in whether the program is successful in improving GCP (outcome evaluation). If the program is a success, DCR is interested in what aspects of the program made it a success (process evaluation)."*

The statement of work also prescribed use of Donald Kirkpatrick's (1975) model for evaluating training programs. According to this model, four levels of training can be evaluated. Information from each prior level serves as the basis for the next level's evaluation, with each successive level representing a more precise measure of training effectiveness.

The four levels of Kirkpatrick's model measure:

- ***Reaction*** to training (thoughts and feelings about the training)
- ***Learning*** of content (increase in knowledge, capacity, skills)
- ***Transfer*** of learned content to trainee's setting (behavior change, use of new skills)
- ***Results/Impact*** of transferred content in trainee's setting (effect on the setting/environment resulting from trainee's performance).

Purpose of Feasibility Study

NIAID's DCR, in partnership with NIAID SPEB, was interested in conducting an evaluation feasibility study of the NIAID GCP CBT program to assess its effects to improve the use of GCP in clinical trial research. DCR expects to use the evaluation findings to justify the value added of the program. If evaluation findings demonstrate effectiveness of GCP CBT, this program may be adopted by other NIH Institutes, including their domestic and international extramural research communities.

The purpose of this NIAID GCP CBT evaluation feasibility study was to determine whether conducting a full-scale evaluation of the program was appropriate, and to identify best possible evaluation designs, methodologies, and data collection strategies to assess the program's effects. The NOVA evaluation team conducted a systematic assessment of optimal plans to evaluate the GCP CBT by proposing questions to be answered by the evaluation, developing data collection instruments, collecting and analyzing data from literature and document reviews, conducting telephone interviews, and identifying appropriate evaluation designs (e.g., pros and cons of evaluation options). This report provides guidelines and specific recommendations to design a full-scale process and outcome evaluation of the NIAID GCP CBT and of GCP compliance.

Evaluation Plan Development

Overview of Evaluation Approach

The development of an evaluation plan incorporates evaluation objectives within a conceptual framework that depicts program activities and outcomes, as viewed by key stakeholders. The NOVA team initiated the evaluation planning process with a face-to-face meeting with NIAID Program

Officers, DCR representatives, and SPEB staff. An Evaluation Planning Group was subsequently formed. Together, they set goals for the evaluation planning process, discussed questions about NIAID's concept of how the GCP CBT program would work, and identified documents that needed to be reviewed. Exhibit 4 describes NIAID GCP CBT goals specified during the feasibility study. The NIAID GCP training represents a step towards consistency and clarity in conduct of NIAID-funded clinical trial research studies following NIAID-specific GCP. Potential users of GCP CBT include clinical investigators, study coordinators, clinical monitors, Program Officers, and others including pharmacists, nurses, case managers, laboratory staff, social workers, and administrative staff. Within the group of potential users, the initial release of this training will target both intramural and extramural researchers and staff.

EXHIBIT 4. NIAID GCP CBT GOALS

The overarching goal of NIAID's GCP CBT is:

To increase knowledge of GCP in conduct of clinical trial research (i.e., best practices in planning, initiation, implementation, closeout, and reporting of clinical trials involving human subjects).

Objectives include demonstrating value added (i.e., improvement) in three areas:

a. Subject safety

b. Data integrity—study data that are more reliable, trusted, accurate, and complete

c. Research efficiency—bringing a research project from the research question to the answer expeditiously.

The underlying assumption of the GCP training is that knowledge acquisition of GCP will facilitate application of GCP in the conduct of clinical trial research. The information obtained through activities of the feasibility study facilitated the identification of two distinct components of the evaluation based on outcomes of interest and realistic attributions of changes in desired outcomes.

Outcomes of interest to be evaluated reflected Kirkpatrick's four-level model (reaction to training, learning of GCP, transfer of GCP to research, and impact of GCP compliance in clinical trial research). However, the extent to which all outcomes could be attributed to GCP CBT differed, leading to the identification of evaluation components described in Exhibit 5.

EXHIBIT 5. EVALUATION COMPONENTS

The two main evaluation components are:

Evaluation Component A: Evaluation of GCP CBT Effects

This component assesses direct effects of the CBT on knowledge acquisition of GCP.

Evaluation Component B: Examination of GCP Compliance

This component focuses on GCP compliance in the conduct of clinical trial research over time.

Evaluation Component A: Evaluation of GCP CBT Effects. This component will evaluate outcomes on reactions to training and learning resulting from GCP CBT training. The actual experience of users with the training (e.g., satisfaction, opinions) and content learned (e.g., GCP knowledge) *as a result of the training* are feasible to be evaluated. An appropriate evaluation design

(see "Answering Evaluation Questions on Component A," subsection 1.4, Goal 3) can facilitate the attribution of any change in GCP knowledge (from pre- to post-test) among trainees taking the GCP CBT courses.

Evaluation Component B: Examination of GCP Compliance[1]. This component will explore change in adherence to GCP (e.g., transfer of GCP knowledge to conduct clinical trial research) and subsequent improvement in subject safety, data quality, and research efficiency (impact of GCP compliance) over time after implementation of the GCP CBT program. However, any change in GCP compliance *cannot be directly attributed* to NIAID's GCP CBT for several reasons:

1. Currently, NIAID's GCP CBT is a *voluntary* training; it is possible that not all staff involved in clinical trial research at a given site participate in the NIAID-sponsored training.
2. GCP CBT learning objectives focus on knowledge acquisition of GCP, *not* on measurable behavioral changes of GCP use. That is, the program has not been designed to directly affect GCP compliance behaviors, but to ensure learning about GCP guidelines. As such, it is only hoped that learned knowledge subsequently facilitates enactment of GCP to conduct clinical trial research. Thus, any increase in GCP use over time can, at best, be only indirectly related to the program (e.g., as when data analysis indicate that an increase in GCP use is more evident in sites where the study's Principal Investigator (PI) and a significant number of research staff have participated in the GCP CBT).
3. Assessing GCP behavioral compliance is complicated by a variety of factors, many of which are unrelated to GCP knowledge but tend to influence the application of GCP in the conduct of clinical trial research.

   As part of the feasibility study, NOVA interviewed site monitoring staff from NIAID Divisions sponsoring clinical trials to identify existing data sources that might be used to evaluate the impact of GCP CBT on GCP compliance. Interviews were conducted with a Health Specialist, a Project Officer for a site monitoring contract, a Clinical Monitoring Coordinator, a Clinical Trials Manager, two Clinical Research Oversight Managers, and a site monitoring contractor from NIAID's Division of AIDS; these individuals represented the Regulatory Compliance and Human Subjects Protection Branch, the Office of Clinical Research Affairs, the Division of Microbiology and Infectious Diseases, and private companies under contract with NIAID to provide oversight of clinical trials site monitoring (see **Appendix H** for summary of interviews). Interviewees concurred that attributing GCP compliance change to the program is challenging because:

   - Many staff are involved in the site monitoring process.
   - Site visits cannot monitor every GCP-associated clinical activity at every visit.
   - Many staff are involved in the conduct of clinical trials (e.g., PI, clinical physicians, study coordinator, nurses).
   - The content of the monitoring process is consistent, but there is variability in aspects such as (a) number of people involved, (b) levels of monitoring because of study characteristics (e.g., design), and (c) issues in conducting research (which are often out of the researchers' control, such as staff changes or delays to initiate a study due to the local

[1] GCP compliance in this report refers to the implementation of NIAID GCP guidelines in the conduct of NIAID-funded clinical trial research studies. The terms GCP *"compliance," "use," "adherence," "implementation,"* and *"application"* are used interchangeably throughout the report, and have the same meaning.

IRB, both of which do not imply an incorrect implementation of GCP) that limit or confound the answers needed.

- Violations can happen because of protocol or study participants. Problems with subjects' behavior and noncompliance issues are likely to affect and skew GCP use data.

4. GCP compliance data will be collected at the research site/study level only; not at an individual level (e.g., each PI, study coordinator, research nurse). An individual's behavioral measures of GCP use and GCP knowledge through CBT cannot be directly linked.

Importantly, the evaluation components complement each other by providing a general picture of NIAID's focus on GCP use in the conduct of clinical trial research. **Evaluation Component A**: Evaluation of GCP CBT Effects will assess changes in GCP knowledge due to a specific NIAID-sponsored training activity—the GCP CBT. **Evaluation Component B**: Examination of GCP Compliance will explore adherence to GCP guidelines in the conduct of clinical trial research over a period of time after GCP CBT implementation.

The GCP CBT training course is being implemented for the first time; thus, learning about its effectiveness to increase GCP knowledge is warranted as a short-term program goal. The expected value-added improvements in subject safety, data integrity, and research efficiency reflect behavioral changes in GCP compliance over time, and as such are long-term goals.

This report is structured to guide the reader in the sequence and framework used to conduct the evaluation feasibility study. The study sequence and framework used allowed the identification of evaluation aspects pertinent to each evaluation component. Findings and recommendations from the feasibility study which are unique to either Evaluation Component A or Evaluation Component B are pointed out throughout the report.

Feasibility Study Questions

NOVA aimed to gain a clear understanding of NIAID's expectations for a full-scale evaluation. The following questions were addressed to inform the evaluation feasibility study:

- What is the purpose and scope of the GCP CBT evaluation?
- What evaluation questions are important to NIAID?
- What practical issues (e.g., methodological, ethical, and financial constraints) need to be considered in planning the evaluation?
- What process and outcome evaluation methodologies and techniques are most appropriate for assessing the GCP CBT program?
- What existing data sources can be used to evaluate this program? What new data need to be collected?
- What comparison groups are available and appropriate?
- What is the cost to collect various types of data in dollars, time commitment, and burden on staff and evaluation participants?
- Is there adequate justification to conduct a full programmatic process and outcome evaluation at this time? If so, what are the most appropriate approaches to use in evaluating program effects?

Questions based on Kirkpatrick's training evaluation model were also addressed and are reflected in both evaluation components:

Evaluation Component A: Evaluation of GCP CBT Effects

*Reaction***:**

- Are metrics and methods available with EduNeering's built-in survey capability adequate to assess whether users perceive the GCP CBT as user friendly, with learning objectives that are relevant and transferable to the workplace?

Learning:

- Are metrics and methods available with EduNeering's pre- and post-testing sufficient to assess knowledge and changes in knowledge of GCP? If not, can metrics and methods be developed and added?

Evaluation Component B: Examination of GCP Compliance

Transfer:

- Are metrics and methods available or can they be developed and easily implemented to measure behavioral changes for those conducting clinical human subject research?

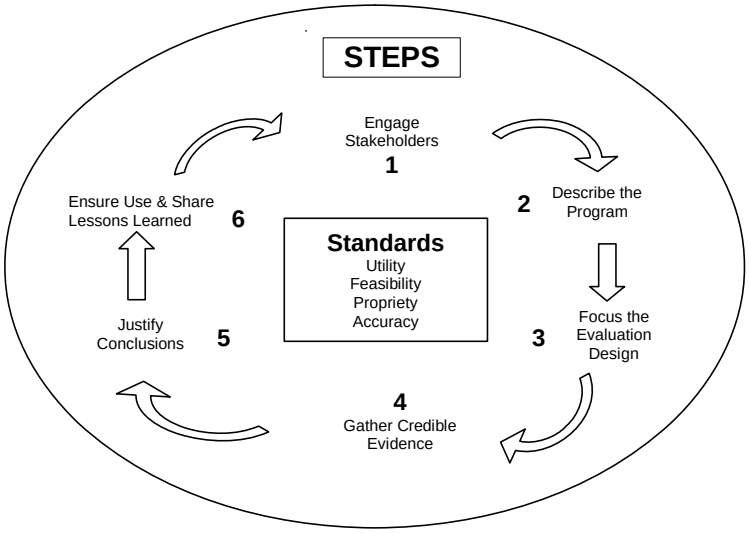
Results/Impact:

- Are metrics and methods available or can they be developed and easily implemented to assess outcomes of clinical human subject research in terms of subject safety, efficiency of completing trials, and confidence in answers derived from the trials (accuracy and completeness of data)?

Theoretical Evaluation Framework

The NOVA team used a framework for program evaluation developed by the Centers for Disease Control and Prevention (CDC). Exhibit 6 shows this "Framework for Program Evaluation in Public Health" (CDC, 1999). NOVA used this framework, which includes a series of steps and tools, for the design of a comprehensive evaluation plan for the NIAID GCP CBT program and examination of GCP compliance in clinical trial research. This step-by-step approach is described to illustrate how it can result in a comprehensive evaluation of GCP CBT and exploration of GCP use that meets the needs of NIAID.

EXHIBIT 6. FRAMEWORK FOR PROGRAM EVALUATION



Source: CDC. (1999). Framework for program evaluation in public health. *Morbidity and Mortality Weekly Report* 48 (RR-11):1-40.

Evaluation Plan

Step 1. Engage Stakeholders

Key stakeholders are defined as individuals or organizations that have an investment ("stake") in what will be learned from an evaluation and what will be done with this information (CDC, 1999; Patton, 1997). Stakeholders are often experts in a program and understand how it can or should affect target audiences or national programs. The stakeholders should include members of the organizations that "bought into" the program.

NIAID GCP CBT program stakeholders include NIAID leadership and executives, staff within NIAID DCR and SPEB, NIAID-funded intramural and extramural researchers and staff, and EduNeering staff. Other stakeholders include contractors and extramural science administrators. NOVA and NIAID program staff identified three principal groups of stakeholders: (1) those involved in GCP CBT program operations (e.g., NIAID program staff), (2) those served or affected by this program (e.g., NIAID intramural and extramural researchers and staff), and (3) primary users of evaluation findings (e.g., DCR and SPEB).

As this evaluation moves forward, other key stakeholders may become involved. The depth of involvement from stakeholders may range from minimal (such as providing feedback on materials)

to extensive (such as completing tasks that have a direct impact on what the program does and accomplishes).

Recommendation:

- **The GCP CBT DCR Project Manager should ensure that the appropriate stakeholders are involved in planning and implementing the evaluation.** Key stakeholders of the evaluation include NIAID leadership and staff within DCR and SPEB. Additional stakeholders may include members of an evaluation advisory committee, other NIAID Divisions in charge of clinical trials (e.g., Division of AIDS—AIDS, Division of Intramural Research—DIR, Division of Microbiology and Infectious Diseases—DMID, and Division of Allergy, Immunology, and Transplantation—DAIT), and individuals from the research community who are conducting clinical trial studies.

Step 2. Describe NIAID GCP CBT Program

The description of a program includes its purpose and information regarding the way it was intended to function and the way it was actually implemented. A clear and accurate description of the program allows for a balanced assessment of its strengths and weaknesses. In addition, it helps stakeholders understand how the program components fit together and relate to the overall goal. A program description includes a delineation of the program theory (i.e., logic model) so that there can be a common understanding of the program's goals, structure, connections, and expected outcomes. The logic model also assists in focusing the evaluation design on the most critical program elements. The evaluation design is then applied to this model.

The NOVA team discussed the goals and objectives of GCP CBT and other issues relevant to the evaluation in early meetings with the NIAID DCR Deputy Director, Project Manager, and Contract Project Officer; SPEB Acting Chief and Project Manager; and other NIAID, EduNeering, and Science Applications International Corporation (SAIC) staff. These meetings facilitated a mutual understanding of program purpose, anticipated activities, resources, stage of development, expected effects, and context.

In addition, the NOVA team reviewed program documents and conducted a literature review of relevant studies on GCP and, specifically, on evaluation of training in the conduct of human subjects research. Recommendations based on the literature review addressing the training program included suggestions for implementation, communication on and about the training, and tracking of training data. Recommendations for evaluation addressed outcome indicators, methodologies for data collection, data analysis and interpretation, and communication of evaluation findings (see **Appendix G** for the approach used in the literature review and recommendations). These activities helped determine the feasibility of an evaluation, identify appropriate models or theories to guide the evaluation (e.g., logic model), and informed the evaluation design (e.g., evaluation components and related outcomes, measures, methodologies, data sources, and analyses).

2.1. Evaluation Program Theory (Logic Model)

Based on input from the Evaluation Planning Group and the activities discussed above, a program logic model was developed to provide a synopsis of the NIAID GCP CBT program. The model shows the relationships among major project aspects, activities and outputs envisioned by the program, and desired outcomes associated with program activities. It provides a logical sequence of the way the resources invested by the GCP CBT program will lead to program refinements and desired results. A logic model generally has the following elements: program inputs, activities,

outputs, short-term outcomes (1-3 years), and long-term outcomes (4-6 years). **Appendix A** shows the logic model developed for the GCP CBT program evaluation.

Recommendation:

- **The program logic model needs to be reviewed annually so that it accurately reflects program changes.** An effective logic model is refined and changed many times throughout the evaluation process as staff and stakeholders learn more about the program, how it operates, and why it works. This process aids in adjusting approaches and changing course as the GCP CBT program evolves over time.

2.2. Outcomes of Interest to the NIAID GCP CBT Program

The evaluation plan focuses on two different but related aspects of the program specified in the logic model: *formative* aspects that reflect implementation of planned program activities (also referred to as *progress* or *process*) and *summative* outcomes that reflect expected short- and long-term program effects (also referred to as *outcomes*).

Common formative or *process* indicators describe program operations and elements of change that are precursors to system changes and contribute to evaluation of summative outcomes (end-of-program). The evaluation of formative aspects assesses the extent to which GCP CBT is being implemented as planned (e.g., dissemination plan) and measured on a regular basis (e.g., quarterly, annually). In addition, the evaluation seeks to measure summative changes brought about by the program in outcomes of interest (e.g., GCP knowledge gain). As mentioned earlier, outcomes of interest identified for the evaluation are in line with Kirkpatrick's model (see the logic model in **Appendix A** for outcomes).

Step 3. Focus the Evaluation Plan

A focused plan increases chances that the evaluation will succeed in providing direction and determining what steps are practical and cost-effective. Among the items to consider when focusing an evaluation are purpose, users, uses, advisory committee, evaluation questions, benchmarks or indicators, and methods.

3.1. Purpose

The purpose of the evaluation is to consistently measure relevance and effectiveness and to produce meaningful reports to interested stakeholders both within and outside NIAID (e.g., other Institutes within NIH). The ultimate goals of the evaluation are to assess the effectiveness of GCP CBT and GCP adherence in the conduct of NIAID-funded clinical trial research studies.

3.2. Users

Key stakeholders should be asked to review and prioritize evaluation questions and methods and the intended uses of the evaluation to prevent it from becoming misguided or irrelevant. To focus the evaluation, NIAID staff and the Evaluation Contractor need to work with other key stakeholders to prioritize areas to address in the evaluation plan. Based on these priorities, feasible evaluation strategies can be refined and integrated into the evaluation plan.

3.3. Uses

The results of this evaluation will be used for multiple purposes, including making appropriate program refinements as information on program operations is gathered regularly; making decisions regarding the continuation of program and funding; and providing lessons learned that can be

applied to training programs similar to GCP CBT. Likewise, information gathered on GCP compliance barriers and facilitators can help identify issues that can be corrected or that need further investigation.

3.4. Advisory Committee

An Evaluation Advisory Committee should be used to evaluate GCP CBT and to assess GCP use. This Committee will serve both technical and practical functions. It will provide expertise and recommendations to focus the scope of the evaluation, methodologies, data collection instruments, and to identify contextual circumstances to consider.

Recommendations:

- **Form an Evaluation Advisory Committee early on to advise on technical and practical functions of the evaluation.** The Committee needs to consist of members with relevant experience. At a minimum, it should have members with experience in evaluation (i.e., evaluation of GCP, evaluation of transfer of learning, adult learning strategy, e-learning strategy, HSP, and other related regulations for conducting clinical trial research), GCP, NIH-related clinical trial regulations, and clinical trial monitoring.
- **Define the Committee's goals and members' responsibilities.** Evaluation Advisory Committee members should be clear about the Committee's purpose, their roles and responsibilities, and the estimated number of meetings per year.

3.5. Evaluation Questions

Evaluation questions establish boundaries by stating what aspects of the program will be assessed. The process of identifying potential information needs often results in more questions than can be addressed in a single evaluation effort. A comprehensive look at potential evaluation questions will make these possibilities clear to the NIAID GCP CBT Evaluation Planning Group, allowing for informed choices when selecting questions.

Both formative (process) and summative (outcome) evaluation questions were developed by the NOVA team for the full-scale evaluation. Answers to formative evaluation questions will provide information that can be shared quickly to improve the program, as these questions focus on program activities, challenges, and outputs for the purpose of monitoring progress and making midcourse corrections, when needed. Answers to summative (outcome) evaluation questions will provide information on short-term and long-term changes. The following formative and summative evaluation questions were identified during the feasibility study; they are separated by evaluation components:

Evaluation Component A: Evaluation of GCP CBT Effects

The information generated will be used to determine the program's overall value and worth, help identify changes necessary to improve accomplishment of overall goals, and determine the extent to which the program has worked as planned and whether it has succeeded as expected.

GCP CBT Awareness and Usage

- Are target audiences aware of the availability of and the way to access the GCP CBT courses?
- What is the rate of completion of GCP CBT courses among target audiences?

Reaction to Training

- What is the overall reaction of trainees to GCP CBT (e.g., satisfaction)?
- How relevant and transferable is GCP CBT to trainees' work setting?

Learning[2]

- What is the influence of GCP CBT on trainees' knowledge GCP?

Evaluation Component B: Examination of GCP Compliance

Data collected will provide information about GCP compliance over time. Information gathered can reveal pressing issues currently hampering adherence to GCP (e.g., across studies, study types, Divisions), those that can be readily corrected or are in need of more in-depth examination, suggestions, and lessons learned in the application of GCP to conduct clinical trial research.

Transfer of GCP Knowledge

- Is there an increase in the application of GCP over time after availability of GCP CBT?
- What are common barriers to and facilitators of GCP compliance?

Results/Impact of GCP Compliance

- Is there an improvement in subject safety over time after availability of GCP CBT?
- Is there an improvement in quality of research data over time after availability of GCP CBT?
- Is there an improvement in research efficiency over time after availability of GCP CBT?

3.6. Benchmarks and Performance Indicators of GCP CBT for Evaluation Component A: Evaluation of GCP CBT Effects[3]

Quality evaluations include assessments that describe the criteria for success. Benchmarks or indicators of program performance are necessary to establish the extent to which the program is accomplishing what it set out to do in terms of process (e.g., outputs) and outcomes (e.g., change in GCP knowledge). Benchmarks and corresponding performance indicators will help answer questions such as: "Is the program moving toward anticipated goals? Do GCP CBT trainees demonstrate adequate knowledge of GCP?" Performance indicators are also useful to monitor ongoing program status against a set of targets (program objectives or goals). Based on set benchmarks, an alert system for unexpected developments or lack of progress can be built into the evaluation plan. This can help key stakeholders (e.g., NIAID program management) review how the program is operating, whether progress is as expected, and whether there are issues or problems that need to be addressed.

Recommendation:

- **Establish realistic benchmarks for process and outcome indicators of GCP CBT success.** Expert opinion of program staff and characteristics of participants often set the benchmarks or indicators of success for each program goal.

---

[2] Due to confidentiality provisions (i.e., personal identifiable information) stated by NIAID's program staff, retention over time cannot be measured.

[3] Benchmarks and performance indicators apply only to Evaluation Component A: *Evaluation of GCP CBT Effects* because set criteria for success are only applicable for the GCP CBT. Evaluation Component B: *Examination of GCP Compliance* is primarily an exploratory assessment of GCP use, so setting criteria for success in adherence to GCP does not apply.

3.7. Timeline

Several factors are considered when determining the nature and scope of a program evaluation. Program characteristics (e.g., intended targets, outcomes of interest, locations), evaluation questions, and other practical considerations (e.g., resources, maturity of program) shape the scale (size, time period) of the evaluation.

GCP CBT is a new program with a purported wide reach (e.g., audiences from various NIAID Divisions and roles, engaging in intramural and extramural clinical trial research, in the U.S. and internationally) and impact. The planned evaluation is comprehensive because of its evaluation components (**Evaluation Component A**: Evaluation of GCP CBT Effects and **Evaluation Component B**: Examination of GCP Compliance), assessment of process and outcomes, short- and long-term goals, and multiple audiences and locations. The evaluation needs to be sensitive to allow, for example, enough time for the GCP CBT to mature so that expected outcomes are measured at the right time. A comprehensive, reliable evaluation is dynamic, expanding and/or switching its focus and activities as the program develops over time to ensure a fair assessment of program effects. A proposed Timeline/Schedule of evaluation activities is provided in **Appendix C.**

Recommendations:

- **Plan for a 5-year evaluation.** A 5-year plan is an appropriate time period to assess both evaluation components. The evaluation should also plan for incremental steps in its scope. For example, the evaluation should focus on intramural research the first 2 years and then assess extramural research; examine domestic clinical trials first and then international studies; and collect longitudinal data over a reasonable period (e.g., 4-5 years) to examine GCP compliance.

Evaluation Component A: Evaluation of GCP CBT Effects

- The GCP CBT program is a new program with a strategic plan to offer training to increased audiences over time (e.g., intramural research first and then extramural; research staff from studies conducted in the U.S. first and then international sites). A 5-year assessment of GCP knowledge change due to CBT among diverse audiences is recommended.

Evaluation Component B: Examination of GCP Compliance

- Due to the diverse nature of clinical trial studies, goals, and related protocols, examination of GCP compliance needs to be assessed long enough to explore behavioral changes and facilitators/barriers to adherence. Collection of longitudinal data on GCP use over 5 years is warranted to assess different clinical trial studies (e.g., intramural, extramural, domestic, and international).

3.8. Evaluation Planning Matrix

An Evaluation Design Matrix is a useful organizational tool that flows from the program Logic Model. It typically includes program evaluation questions, type of data needed to answer evaluation questions, methods to be used, data sources, data analysis, limitations of findings, and knowledge gained from the evaluation. In collaboration with the Evaluation Planning Group, the NOVA team created a matrix that describes how evaluation questions pertaining to each component will be answered (see **Appendix B**).

3.9. Methods

A mixed-method approach using quantitative and qualitative evaluation measures strengthens the overall evaluation design by allowing for more precise statistical measurement in the quantitative components and in-depth insight in the qualitative components. For example, for **Evaluation Component A**: Evaluation of GCP CBT Effects, quantitative data can produce estimates of the prevalence of knowledge, opinions, and other characteristics of GCP CBT-target audiences. For **Evaluation Component B**: Examination of GCP Compliance, quantitative data can produce estimates of adherence to study protocols, IRB compliance, and safety violations. Surveys are commonly used to collect quantitative data and are administered in a variety of ways (e.g., in-person, telephone, Internet). Qualitative data helps to produce a better understanding of issues and evaluation findings. Techniques for gathering qualitative information vary greatly (e.g., in-depth individual interviews, focus groups). Both quantitative and qualitative data will be collected to answer evaluation questions for each component.

Recommendations:

Evaluation Component A: Evaluation of GCP CBT Effects

- **Use quantitative and qualitative data collection methods.** Use a mixed-method evaluation design of qualitative and quantitative techniques to obtain a full picture of GCP CBT effects.
- **Ensure appropriate sample sizes to determine statistical significance of quantitative results.** The traditional approach to determine program effects based on quantitative data analyses is statistical significance. Statistical significance is compromised by small samples. The use of power analysis (for effect size) that specifies sample size to determine statistically significant changes (e.g., improved GCP knowledge from pre- to post-test) is recommended (see "Answering Evaluation Questions to Evaluation Component A," Goal 3, *Recommendations*).
- **Minimize burden on respondents and ensure confidentiality or anonymity of sensitive personal information**. Data from tests can be obtained unobtrusively through system devices that ensure data integrity and meet DHHS/NIH requirements for data confidentiality and security.

Evaluation Component B: Examination of GCP Compliance[4]

- **Use quantitative and qualitative data collection methods.** Use a mixed-method evaluation design of qualitative and quantitative techniques to obtain a more comprehensive depiction of GCP compliance, including barriers and facilitators.
- **Minimize burden on respondents and ensure confidentiality or anonymity of sensitive personal information**. Data from surveys, interviews, and focus groups on clinical trial studies can be obtained unobtrusively through system devices that ensure data integrity and meet DHHS/NIH requirements for data confidentiality and security.

[4] Provisions for statistical significance for Evaluation Component B: Examination of GCP Compliance are not applicable for a variety of reasons: data will be collected only on new studies (the number of funded studies is unknown), a wide range of clinical trial studies will need to be accounted for, and the level of participation of study sites in the evaluation is unknown.

Step 4. Gather Credible Evidence

Collecting data through a mixed-method approach from various key sources that address program implementation and outcomes makes it easier to collect credible evidence. Different types of data obtained from various sources and by different methods will convey a well-rounded picture of the program and a general overview of compliance with GCP; these data will be used by the evaluation's primary users to draw conclusions and make informed decisions. Although all types of data have limitations, multiple procedures (e.g., qualitative, quantitative, surveys, interviews) for gathering, analyzing, and interpreting data enhance the quality of the data (CDC, 1999). The collected information will present a clear and reliable picture of NIAID GCP CBT program effects and an overview of GCP adherence during the conduct of clinical trial research after GCP CBT implementation.

4.1 Data Sources

A variety of sources can provide the necessary information for the evaluation. Program records, NIAID GCP CBT program staff, NIAID non-program staff, target audiences, and key informants are some sources of data that will be used for a comprehensive program evaluation. Next is a description of the data that will be collected. The way(s) in which these data will be used to answer evaluation questions is outlined in the Evaluation Matrix (see **Appendix B**). Final versions of data collection instruments will be developed and implemented with input from the Evaluation Advisory Committee and pertinent audiences (e.g., site monitors, program staff). Data sources will be explained for each evaluation component.

Evaluation Component A: Evaluation of GCP CBT Effects

4.1.1. EduNeering Database

The NIAID GCP CBT system database developed by EduNeering will collect data on GCP CBT program aspects, including: 1) NIAID's User Profile (nature of working relationship with NIAID—e.g., employee, contractor, grantee, collaborator, Division association within NIAID, role in NIAID-sponsored clinical trials, career stage, and previous experience with GCP and HSP and related certifications; see **Appendix D**); 2) number of participants; 3) course completion; and 4) pre- and post-knowledge test scores. Data for the evaluation will be aggregated to protect confidentiality and anonymity of participants.

Recommendations: If possible, it is suggested that EduNeering incorporate the following data elements and related procedures into the database:

- **Include item-level responses by user ID**. Collecting data by user ID for both pre- and post-knowledge tests helps better assess changes in knowledge.
- **Collect descriptive metrics on GCP post-test retake**: number of people retesting, number of times retesting, test scores at each test taking (for course, module, and "challenge" post-test) to obtain additional information on weaknesses and strengths in GCP knowledge.
- **Collect CBT participation data.** These data (number of users who access/open the CBT Web site; number of users who access at least one course of GCP CBT; courses most frequently accessed, started, and completed by users) will help assess awareness, access, and most common deficiencies in GCP knowledge among users.

4.1.2. GCP CBT User Satisfaction Survey

A brief survey that assesses participant's overall experience with the training will be administered at the conclusion of each CBT training session. Survey items were developed as indicators of Kirkpatrick's four-component model (i.e., reaction, learning, transfer, and results) and to answer related evaluation questions. Questions address level of satisfaction, perceived usefulness of training to enhance knowledge of GCP, and relevance of the training to the workplace. The survey will be optional and will be administered at the end of the post-test (see **Appendix E**).

4.1.3. Focus Groups with GCP CBT Users

Focus group discussions allow for in-depth probing of pertinent topics. The purpose of conducting focus groups with GCP CBT users is to gain a more thorough understanding of critical issues regarding the CBT based on results from the User Satisfaction Survey and users' comments offered in the *helpdesk* option available at the end of the training. At least two types of focus groups should be conducted based on participants' roles or positions (and therefore experience with GCP) to get data from different perspectives on issues and where they intersect. One type of focus group would consist of staff who provide project oversight of clinical trials (e.g., Site monitors, Program officers); the other type would consist of research staff who conduct clinical trials (e.g., PI, study coordinator). Data from focus groups and the Satisfaction Survey will help identify successes, processes that led to these successes, and recommendations to improve the GCP CBT program.

Evaluation Component B: Examination of GCP Compliance[5]

4.1.4. Site Monitoring Tool

As mentioned earlier, the NOVA team interviewed site monitoring staff from NIAID Divisions that sponsor clinical trials to identify existing data sources that might be used to examine compliance with GCP. Measuring GCP use in the site monitoring process is challenging for a variety of reasons—for example, multiple entities involved in the site monitoring process, only certain GCP aspects monitored each visit, large number and functional diversity of research staff. Interviews helped identify "core" or common key elements that are regularly monitored and collected across Divisions (see **Appendix H** for a summary of interviews). These elements will be incorporated in a Site Monitoring Tool to be used by site monitors to collect data on GCP use and compliance. This tool will also collect site monitors' general perceptions of GCP use at a given site. A final version of the tool will be developed and implemented with input from site monitors (e.g., via pilot focus groups) to ensure that information on key elements and response options are appropriately collected.

4.1.5. Survey on Issues with GCP Compliance in Conduct of Clinical Trials

Information on facilitators and barriers to GCP compliance in NIAID-funded clinical trial research will be obtained. Areas to be examined will be based on core elements from the Site Monitoring Tool regarding application of GCP guidelines, including barriers and promoters in the adoption of regulations, protocol approval (e.g., Informed Consent Forms, IRB process), and participant recruitment. Relevant information gathered from the Site Monitoring Tool will also be incorporated into the survey. Survey questions will address experiences with GCP processes and ongoing issues

[5] Data from study-related archival documents to obtain information on GCP compliance in the past will not be collected because in discussions with the Evaluation Planning Group and interviews with NIAID site monitoring staff (see Appendix H), it became clear that accessing these documents (e.g.., site monitoring reports) is complicated due to ethical concerns and issues with confidentiality on any specific study.

and will be completed by project oversight officials from NIAID monitoring staff and independent monitoring contractors and PIs of clinical research studies.

4.1.6. Focus Groups with Project Oversight and Grantee Staff of NIAID-Funded Clinical Trials

The purpose of conducting these focus groups is to gain in-depth understanding of key factors, including those beyond knowledge of GCP (e.g., through CBT), that are frequent barriers to GCP compliance in the conduct of clinical trial research. Two types of focus groups are recommended according to participants' roles or positions to obtain diverse perspectives on issues. One type of focus group would consist of staff providing project oversight of clinical trials (e.g., Site monitors, Program officers); the other type would consist of research staff conducting clinical trials (e.g., PI, study coordinator). Some of the findings from the Survey on Issues Related to GCP compliance will also be addressed in the focus groups for further clarification and understanding. Data from the focus groups and Survey on Issues Related to GCP compliance will help identify common issues, whether barriers are likely to vary by type of study, and lessons learned that can be recommended for new studies.

4.1.7. Interviews with NIAID Division Representatives

Interviews with NIAID Division representatives should be conducted to explore perceptions regarding practices across Divisions that can be associated with NIAID GCP and the CBT. Representatives from NIAID Divisions directly involved in the conduct of clinical trials will be interviewed (e.g., DAIDS, DIR, DCR, DAIT, DMID) about the applicability of NIAID GCP to all or most research studies, its perceived impact on clinical research, its and influence on working relationships across Divisions.

4.2. Data Analysis

4.2.1. Quantitative Methods

Analysis of quantitative measures will be addressed for each evaluation component.

Evaluation Component A: Evaluation of GCP CBT Effects

Analysis of quantitative measures will begin with descriptive statistics (e.g., frequencies, means, cross-tabs) to characterize the data and answer evaluation questions related to the GCP CBT program. Other quantitative analysis will focus on aspects related to training participation and content and comparison of pre and post scores from the GCP knowledge test developed for the training. These data should be examined quarterly to monitor progress and detect any problems that require intervention. More complex analyses and causal modeling, such as analyses of variance and regression analysis, may be possible depending on the quality and quantity of the data. If the data support these more complex analyses, these would be performed as part of the summative evaluation (short-term outcomes).

Evaluation Component B: Examination of GCP Compliance

Data analysis on adherence with GCP guidelines should be performed annually or every 2 years, based on decisions by NIAID, and would be primarily descriptive in nature. Descriptive data from other sources, such as the Survey on Issues Related to GCP Compliance, should be used to complement understanding and measurement of GCP use.

4.2.2. Qualitative Methods

The approach to analysis of qualitative data is similar for both evaluation components.

Qualitative data from focus groups and interviews should be transcribed verbatim. These data should be analyzed and interpreted using content analysis in which main ideas, comments, and words are grouped based on variables of interest (Patton, 2001). To maximize reliability, coding (i.e., categorizing) of data and thematic analysis of text should be conducted by a minimum of two experienced evaluators. Qualitative software, such as ATLAS.ti, should be used for these analyses.

Step 5. Justify Conclusions

For both evaluation components, the NIAID Program Director and key stakeholders should work together with the evaluator to determine interpretations and conclusions supported by the evidence gathered.

Evaluation Component A: Evaluation of GCP CBT Effects

Findings from the evaluation of GCP CBT effects should be judged against desired outcomes (benchmarks) identified by key program stakeholders. Conclusions on the basis of evidence gathered and analyzed include comparison of program objectives (predetermined measures of success) with analysis and synthesis of information, interpretation of evidence, and recommendations for consideration (CDC, 1999; Patton 1997). When appropriate, conclusions will be strengthened by: (1) summarizing plausible mechanisms of change (e.g., participation in GCP CBT led to knowledge of NIAID GCP and its likely use in clinical trials); (2) delineating temporal sequences between activities (e.g., training) and effects (e.g., increased GCP knowledge); and (3) showing that program effects can be replicated (CDC, 1999; Patton 1997).

Evaluation Component B: Examination of GCP Compliance

Findings from the examination of GCP compliance need to be looked at with caution because of the variety of influences on adherence. Conclusions based on findings will be more useful if they are organized by type of studies (e.g., R34s) and whether issues with GCP compliance are more likely to be internal (e.g., lack of GCP knowledge among most research staff) or external (e.g., infrastructure of academic institution where clinical trial takes place).

Step 6. Ensure Use and Sharing of Lessons Learned

The evaluation process assumes that stakeholders are aware of the evaluation's overall design (e.g., goals, procedures, methods), implementation, and findings to facilitate use of findings when implementing decisions or actions that affect the program (evaluation findings provide a rationale for decisions). Evaluation process activities include designing the evaluation to answer evaluation questions; providing feedback to stakeholders regarding interim findings; and disseminating to stakeholders procedures used and lessons learned from the evaluation (Patton, 1997).

Annual and summative evaluation reports would be submitted to NIAID. Feedback from stakeholders and other users of this evaluation is necessary to ensure that findings are applied. As directed by the NIAID Program Manager, dissemination of lessons learned may include support for writing manuscripts, preparing presentations (e.g., content, slides, handouts), or developing other tailored communication strategies to meet the needs of stakeholders.

Answering Evaluation Questions on Component A: Evaluation of GCP CBT Effects

In order to answer GCP CBT program evaluation questions on awareness, learning, and use of GCP CBT, and reaction to training, specific information on key variables is needed. Data on key variables will be used to develop and interpret findings, and to prepare recommendations and lessons learned specific to the GCP CBT program.

Key Variables

Information on key variables to answer GCP CBT program evaluation questions include program resources, population characteristics, program activities, external factors, and program goals.

1.1. Program Resources

This information includes the amount of funding, human capital, infrastructure, and other resources allocated to the program. GCP CBT program resources include NIAID GCP CBT funding, NIAID staff, the NIAID Clinical Research Subcommittee (NCRS), and NIAID contractors. Data on these variables will need to be provided by NIAID GCP CBT program staff to the evaluation contractor.

1.2. Population Characteristics

These variables describe the characteristics (e.g., demographics) of program participants which may be related to the success of the program. GCP CBT population variables include type of trainee (i.e., intramural, extramural researchers or Clinical Research Organization-CRO-staff, NIAID program officials and site monitoring staff, Division regulatory staff), level of trainee (e.g., PI, study coordinator, research nurse, pharmacist), location of trainee (i.e., international or U.S.), and experience with GCP or HSP (e.g., years of experience, prior certifications). These data will be available through the EduNeering database and NIAID's User Profile Survey (see **Appendix D**).

1.3. Program Activities

These variables depict program operations, processes, or other critical activities. Related variables for the program are GCP CBT dissemination or promotion plans and program modifications. These are primarily process data that can be collected through forms that gather data on planned activities (e.g., dissemination activities to inform new audiences), whether objectives were met (e.g., number and type of activities that were conducted), barriers found during implementation, decisions to address barriers, program modifications, and lessons learned in the process. These data will need to be provided by NIAID GCP CBT staff to the evaluation contractor on a regular basis (e.g., quarterly).

1.4. External Factors

External factors are conditions or circumstances beyond the control of the program which may influence the program's success. These variables provide a context for interpreting the data gathered throughout the evaluation. GCP CBT program variables include problems encountered during implementation of the training (decisions made to address them, lessons learned); perceived reasons for the success or lack of success of the program; NIAID Clinical Research Subcommittee (NCRS) consideration of mandatory GCP CBT, and unexpected positive and negative events occurring during the period under examination. Forms used to monitor program activities will also collect these data and will need to be provided by NIAID GCP CBT program staff to the evaluation contractor on a regular basis (e.g., quarterly).

1.5. Program Goals, Performance Measures, and Comparison Measures

These variables are interrelated and focus on the program's outputs (from implementation activities) and outcomes (effects on target audiences). To the extent possible, each program goal is associated with performance and comparison measures. The overarching GCP CBT program goals mentioned earlier were detailed for the evaluation and reflect short-term program goals (see **Appendix A,** GCP CBT Logic Model). A description of how detailed goals can be evaluated, including options and implications and pros and cons of alternatives, will be discussed where appropriate.

Goal 1: Target audiences are aware of GCP CBT
(Short-term process goal)

The accomplishment of this program goal can be evaluated by examining:

- The GCP CBT comprehensive dissemination plan to inform target audiences of the availability of the training (e.g., type of dissemination activities, message content).
- Whether there is an increased use of GCP CBT over time by target audiences.

Performance Measures (outputs)*:* Tracking sheet on dissemination activities from NIAID program staff (e.g., number and type of activities, targets of disseminations activities), periodic (e.g., annual) cross-sectional surveys (e.g., awareness of available GCP CBT), and CBT use log (e.g., number of users over time, users' characteristics).

Comparison Measures (outputs)*:* Changes over time as compared with the original plan in the tracking sheet on dissemination activities (e.g., number and type of activities, targets of disseminations activities), periodic (e.g., annual) cross-sectional surveys (e.g., awareness of available GCP CBT), and CBT use log (e.g., number of users over time, users' characteristics).

Recommendation:

- **Create a detailed dissemination plan to better assess program implementation.** The dissemination plan should be specific, describing the number of expected activities directed to whom, by whom, and by when, and identifying benchmarks for activities (e.g., minimum number of activities by type planned to implement). The evaluation will assess the extent to which the dissemination plan was implemented as planned, barriers, corrective courses of action, and lessons learned in the process.

Goal 2: GCP CBT users have favorable reactions to the training
(Short-term outcome goal, *Reaction to training*)

The accomplishment of this program goal can be evaluated by examining whether:

- GCP CBT trainees are satisfied with the training.
- GCP CBT trainees perceive that the training is relevant to their job tasks.

Performance Measures: Average ratings from the GCP CBT User Satisfaction Survey of trainee interest, satisfaction, and perception of relevance and transferability of GCP CBT administered at the end of the training. This will be an optional survey (see **Appendix E**).

Comparison Measures: Set criteria for acceptable ratings from the GCP CBT User Satisfaction Survey of trainee interest, satisfaction, and perception of relevance and transferability of GCP CBT.

Recommendation:

- **Set benchmarks for expected ratings on overall satisfaction with CBT.** Set criteria should specify (a) the minimum rating score that would demonstrate CBT satisfaction and (b) the proportion of users expected to score at least the minimum satisfaction score.

Implications of Evaluation Design

- Since it is an optional survey, it is likely that only a small number of GCP CBT users complete it and that results are biased based on characteristics of those who respond.
- Survey findings may indicate the need to make significant changes in the training, ranging from content to system design.

Recommendations:

- **Conduct a pilot study.** A pilot of GCP CBT, including administration of evaluation instruments (e.g., satisfaction survey) and focus groups, will ensure that issues of relevance to the program are addressed before the training is offered widely.
- **Collect data that will be used to improve the program.** Survey items should collect information that will be of use to NIAID program staff and stakeholders to make program modifications, if necessary.

Goal 3: GCP CBT increases knowledge of GCP[6]
(Short-term outcome goal, *Learning*)

The accomplishment of this program goal can be evaluated by examining whether:

- GCP CBT trainees demonstrate higher levels of knowledge about GCP at the end of the training than before the training.
- GCP CBT trainees demonstrate higher levels of knowledge about GCP than target audiences who do not take GCP CBT.

Evaluation Design Options

In program evaluation, an **experimental design** in which participants are randomly assigned to program and control groups provides the highest degree of confidence for drawing causal inferences in outcomes (e.g., GCP CBT is an effective program to increase knowledge of GCP in target audiences). A control group consists of members from the program target audience that had the same chance as program participants to receive the program but, as a result of random assignment, were assigned to a group to which the program was not delivered (at least temporarily). The idea is to have groups that are equivalent to eliminate bias in the results based on members' characteristics. It follows that a larger difference in GCP knowledge scores (from pre- to post-test) accomplished by the program group compared with the control group indicates that the CBT is most likely the cause for the increase. Performance and comparison measures for this design would involve:

[6] EduNeering is in charge of developing the GCP knowledge tests to be used in the training (e.g., knowledge learned on a given topic, overall pre- and post-tests) in collaboration with program staff. Pre- and post-tests on GCP knowledge will be used for the evaluation. It was recommended that tests include items using case scenarios in which users are asked about the correct application of GCP.

Performance Measures: Pre- and post-tests with CBT trainees to determine whether they are more knowledgeable about GCP and better apply GCP guidelines to case scenarios after taking the training than before.

Comparison Measures: Pre- and post-tests with a group of GCP CBT program target audiences who have *not* taken GCP CBT (control or comparison groups).

It is not always feasible to use an experimental design because of practical, ethical, legal, and cost issues. The next best approach to an experimental design is a **quasi-experimental design** in which random assignment to program and control groups is not an option, but which can still allow for scientifically rigorous assessment of outcomes. Regular discussions with the Evaluation Planning Group indicated that a quasi-experimental evaluation design was a more viable option to assess GCP CBT program outcomes.

Two main factors will influence the quasi-experimental design to be implemented: (a) pretest requirements (mandatory vs optional) and (b) existence of a comparison group[7]. There are benefits and drawbacks for the internal validity of evaluation findings based on decisions about pretest requirements and use of a comparison group. Currently, GCP CBT is a voluntary training and as a result, two types of comparison groups are suitable for the evaluation:

(a) Comparison group 1—those who failed the pretest and did not complete GCP CBT
(b) Comparison group 2—those who have neither taken the pretest nor any GCP CBT

Specific criteria need to be set for Comparison group 1 on minimum exposure to GCP CBT to be considered an appropriate "comparison group" (e.g., did not take any CBT or completed less than one-third of failed topics).

*Note: All potential GCP CBT users have the opportunity to test out of the training if they demonstrate GCP knowledge by obtaining a 100% score on the pretest. Those who test out of the training will not be part of the evaluation. Those who obtain less than a 100% pretest score are directed to course topics not passed. All users who take any GCP CBT course are required to take a post-test at completion of training (mandatory to certify training completion.)

Discussion of pros and cons of proposed evaluation designs focus on CBT effects on GCP knowledge[8]. This discussion will be based on the target populations for evaluation: (1) CBT users only versus (2) CBT users *and* a comparison group[9].

(1) GCP CBT target users as sole evaluation population

Next is a description of evaluation designs that consider users of GCP CBT as the *only* population for the evaluation (no comparison group). Table 1 displays evaluation designs that will be addressed.

---

[7] A comparison group is an alternative to a "control group" when the latter is not possible (e.g., random assignment). A comparison group is similar to a control group in that the comparison group does not receive (or minimally receives) the program under evaluation (CBT).

[8] It is likely that many in the GCP CBT target audience have been exposed to some form of GCP training (e.g., HSP) before the program. Previous exposure to or knowledge of GCP will be controlled for in the analysis. This information is captured through the User Profile (see Appendix D).

[9] Discussions will not include all potential designs or pros and cons of a given design, but will describe those considered "feasible" and "appropriate" to assess GCP CBT program effects.

The discussion of each design will focus on implications (pros and cons) of an optional versus mandatory pretest and assumes that post-test is mandatory.

Table 1. GCP CBT Target Users as Sole Evaluation Population (Post-test is mandatory)	
Design A	**Pretest**: Optional/user takes it
Design B	**Pretest**: Optional/user does *not* take it
Design C	**Pretest**: Mandatory

Design A

Pretest: Optional and user takes it **Post-test**: Mandatory

Pros

Results will show tell if there was any knowledge change from pre- to post-test.

Possible bias due to unique characteristics of group who choose to take the pretest can be minimized during analysis (controlling effects of differences at pretest).

Cons

Possible bias due to unique characteristics of this group which may influence their learning of GCP. Any knowledge change found cannot be generalized to other members of the target audience.

Few users may choose to take the pretest, resulting in a small sample size to make inferences of knowledge change (low power of effects for the analysis).

Will NOT know if those *not* taking the pre-test also increased their GCP knowledge.

Will NOT know if any change in knowledge is related to CBT or some other plausible explanation (e.g., an organized GCP seminar).

Design B
Pretest: Optional and user does *not* take it **Post-test**: Mandatory
Pros Decreased user burden. **Cons** Results will show only the amount of user's knowledge after taking all CBT (post-test score). There is no parameter (pretest score) to determine whether any change in knowledge actually occurred (increase). Will NOT know if user's post-test score (whether low or high) is related to taking GCP CBT.

Design C
Pretest: Mandatory **Post-test**: Mandatory
Pros Results will show any knowledge change from pre- to post-test among all CBT users. **Cons** Will NOT know if any change in knowledge is related to CBT or another plausible explanation (e.g., an organized GCP seminar).

(2) GCP CBT target users and comparison groups as evaluation populations

A description of evaluation designs that consider users of GCP CBT *and* comparison groups as populations for the evaluation will be provided. Table 2 displays evaluation designs that will be addressed. The discussion of evaluation designs and implications varies based on type of comparison group and completion of *optional* post-tests.

*Note: The biases introduced by optional tests discussed above also apply to these evaluation designs.

Table 2. GCP CBT Target Users and Comparison Groups as Evaluation Populations

Comparison Group 1 (Those who failed the pretest and did not complete CBT)		**Comparison Group 2** (Those who took neither the pretest nor any CBT)	
Design D	**Pretest**: Available (optional or mandatory) **Post-test**: Available (optional)	Design F	**Pretest**: Available (optional or mandatory) **Post-test**: Available (optional)
Design E	**Pretest**: Available (optional or mandatory) **Post-test**: Not Available	Design G	**Pretest**: Not Available **Post-test**: Available (optional)

Comparison Group 1

Comparison Group 1 consists of members of the target audience who took the pretest (whether optional or mandatory), but did *not* complete the CBT. These persons can be contacted for the evaluation and asked to complete the post-test only. Evaluation Designs D and E consider CBT users and Comparison Group 1 as evaluation populations.

Design D

Pretest: Available (*optional or mandatory)* **Post-test**: Available (*optional)*

Pros

Possible bias due to unique characteristics of this group can be minimized during analysis (controlling effects of differences at pretest).

Will know if there was any knowledge change from pre- to post-test scores between CBT users compared with members of comparison group 1. Differences seen among CBT users, and not the comparison group, can be attributed to CBT.

Helps control for test effects (e.g., better post-test scores due to users' familiarity with pretest).

Cons

Possible bias due to unique characteristics of this comparison group (e.g., those who failed the pre-test and chose NOT to complete CBT). They may not be comparable to those who completed CBT.

Possible contamination of results due to some exposure to GCP CBT.

Design E

Pretest: Available (*optional or mandatory)* **Post-test**: Not Available

Pros

Data are already available (pretest scores).

Minimum burden on participants.

Cons

Possible bias due to unique characteristics of this comparison group.

Possible contamination of results due to some exposure to GCP CBT.

Pretest scores of members of the comparison group will be compared with users' post-test scores on *similar* items only.

Comparison Group 2

This comparison group consists of members of the target audience who *neither* took the pretest nor completed the CBT. These persons can be contacted for the evaluation and asked to complete the post-test only or both (pre- and post-tests). Evaluation Designs F and G which will be discussed consider CBT users and comparison group 2 as evaluation populations.

Design F

Pretest: Available (*optional or mandatory)* **Post-test**: Available (*optional)*

Pros

Possible bias due to unique characteristics of this group can be minimized during analysis (controlling effects of pretest differences).

Will know if there was any knowledge change from pre- to post-test scores between CBT users compared with members of this comparison group. Differences seen among CBT users, and not the comparison group, can be attributed to CBT.

Helps control for test effects (e.g., better post-test scores due to users' familiarity with pretest).

Reduced burden on participants compared with CBT users, as only pre- and post-tests will be administered.

Cons

Possible bias due to unique characteristics of this comparison group (they may not be similar to CBT users).

Design G	
Pretest: Not Available	**Post-test**: Available (*optional)*
Pros	
Minimum burden on participants.	
Results will show if one group is more knowledgeable about GCP than the other.	
More and similar items (post-test) will be used for comparison purposes.	
Cons	
Possible bias due to unique characteristics of this comparison group.	
It will not be possible to determine whether difference in scores is because of CBT.	

Recommendations:

- **Minimize "cons" to evaluation design.** Decisions regarding which evaluation design to implement should minimize threats to validity of evaluation findings (cons).
- **Use a comparison group to assess knowledge change.** GCP CBT is a new program that, if successful, may be modeled across NIH Institutes and Divisions engaged in clinical trial research. It is important to assess its effectiveness to increase GCP knowledge. An evaluation design that includes a comparison group with pre- and post-tests can best attribute differences in GCP knowledge to CBT.
- **Use a comparison group with no or very minimal exposure to the program**. This will guarantee that observed changes between the GCP CBT program group and the comparison group are most likely due to the CBT program. Comparison groups 1 and 2 can be combined to increase needed sample size to achieve statistical significance (see recommendation below on sample size).
- **If a comparison group is not possible, the evaluation should be based on mandatory pretest and post-test.** Mandatory tests will reduce bias of self-selection. To minimize users' burden, the pretest can be mandatory temporarily (until an adequate sample size is accomplished to demonstrate statistically significant effects).
- **Use appropriate sample size to demonstrate program effects (effect size).** Sample size can either be predetermined using a power analysis calculation with an estimated expected percentage change in knowledge or a percentage of participants expected to pass the knowledge post-test. Using power analysis, expected large changes in scores from pre- to post-tests often require a smaller sample size than if a small percentage change is expected. Expected percentage changes are determined by the expertise of program staff and the characteristics of the target audience (e.g., prior knowledge of GCP may lead to a small increase in knowledge from pre- to post-test). For example, assuming a power of .80 and alpha=.05, the minimum number of participants needed to demonstrate a 10% change in scores from pre- to post-test (effect size) is 126; for a 20% change, it is 61 participants; and for a 30% change, it is 56 participants.

Individual Data versus Aggregate Data

Another consideration for the GCP CBT evaluation is how the data will be collected for analysis. Data can be collected at an individual or aggregate level. At an individual level, a participant's data are linked to a nonidentifying unique ID to protect the person's anonymity and the confidentiality of his or her answers. There is no need for a nonidentifying ID linked to a participant's responses if data are collected and reported only in an aggregated state. EduNeering's capacity and collaboration to collect data as suggested will determine the outcome of this recommendation.

Recommendation:

- **Collect data at an individual level.** Matching a person's pretest score to the person's post-test score allows for a more precise analysis of change in knowledge. Analysis of individual data is based on those who took both the pre- and post-test. When data are collected in an aggregated manner, corresponding analysis includes all available pre- and post-test scores indiscriminately. As a result, it is possible that some users who took the pretest do not take the post-test (e.g., scored 100% on pretest) and similarly, some users only take the post-test—if the pretest is not mandatory—introducing bias in evaluation findings.

Answering Evaluation Questions on Component B: Examination of GCP Compliance

In order to answer evaluation questions on GCP compliance—transfer of GCP knowledge to clinical trial research and value added in subject safety, data quality, and research efficiency—data on related key variables are needed to develop and interpret findings, and prepare recommendations and lessons learned specific to GCP compliance.

Key Variables

Due to the nature of Evaluation Component B, in which findings on GCP compliance cannot be directly attributed to a specific NIAID initiative, regulation, or program (e.g., NIAID's GCP CBT), key variables only concern expected long-term goals.

Program Goals, Performance Measures, and Comparison Measures

A description of how GCP compliance goals can be examined, including options and implications and pros and cons of alternatives, will be discussed.

Goal 4: Increased GCP compliance in conduct of clinical trial research
(Long-term goal, *Transfer of GCP knowledge*)

Accomplishment of this goal can be evaluated by examining whether:

- Information on core GCP elements shows a trend toward increased GCP use over time.
- Site monitors perceive increased application of GCP over time.
- Perceptions of barriers to GCP compliance decrease over time.

Performance Measures: Data from the Site Monitoring Tool completed by site monitors and from the Survey on Issues with GCP Compliance completed by NIAID project oversight staff (e.g., program officers), independent monitoring subcontractors, and study PIs in a new site before GCP

CBT; and focus groups with NIAID oversight staff, site monitors, and PIs on barriers and facilitators of GCP compliance and lessons learned conducted during Year 2 of the evaluation.

Comparison Measures: Data from the Site Monitoring Tool reviewed annually over the course of the evaluation; data from the Survey on Issues with GCP Compliance completed during the first 2 years of a study and every other year over a maximum of three and a minimum of one data collection points; and focus groups conducted in Year 4 of the evaluation.

Goal 5: Improved subject safety
(Long-term goal, *Results of GCP compliance*)

The accomplishment of this goal can be evaluated by examining whether:

- Existing data show a trend toward increased subject safety over time.

Performance Measures: Related data from the Site Monitoring Tool completed by site monitors for the first time 6 months after program award.

Comparison Measures: Related data from the Site Monitoring Tool completed by site monitors annually over the course of the evaluation.

Goal 6: Improved quality and integrity of clinical research data
(Long-term goal, *Results of GCP compliance*)

The accomplishment of this goal can be evaluated by examining whether:

- Existing data show a trend toward increased understanding of procedures to ensure high-quality research data over time.

Performance Measures: Related data from the Site Monitoring Tool completed by site monitors for the first time 6 months after program award.

Comparison Measures: Related data from the Site Monitoring Tool completed by site monitors annually over the course of the evaluation.

Goal 7: Improve research efficiency.
(Long-term goal, *Results of GCP compliance*)

The accomplishment of this goal can be evaluated by examining whether

- NIAID Division representatives, Program officials, and monitors perceive an improvement in quality and pace of clinical trials performance/completion.
- Existing data show a trend toward increased GCP compliance over time.

Performance Measures: Interviews with NIAID representatives involved in the conduct of clinical trials and focus groups with NIAID oversight staff, site monitors, and PIs conducted during Year 5 of the evaluation; and summarized data from the Site Monitoring Tool over the 5-year period.

Recommendations:

- **Collect data on GCP compliance repeatedly over the course of the evaluation.** Repeated-times data collected in adequate time spans is necessary to examine any change in GCP compliance within and across studies. If data such as that collected with the Site Monitoring Tool cannot be collected annually, at least baseline data can

be collected in Year 1 before CBT, at the end of Year 3, and at the end of Year 4 to better assess any change (increase) in GCP use.

- **Collect baseline data before GCP CBT is widely implemented.** This will help to better assess program changes attributable to the training. The potential influence of the program on any change in GCP compliance before the training compared with after the training is enhanced as a plausible explanation.
- **Consider a "late baseline" (after CBT started) as another option.** If baseline data collection is not possible before GCP CBT is implemented, an alternative is a "late baseline." In this case, baseline data should be collected soon after the program begins. Exposure to GCP CBT (although minor if collected soon after program startup) will need to be addressed when interpreting evaluation findings.
- **Collect GCP compliance data on newly initiated clinical trial studies only.** This will help better control and assess issues with GCP compliance since program startup and identify barriers across studies as they move along (e.g., types of issues that occur at given stages of the study). Moreover, examining new studies will highlight the role of the program in any change in GCP use. Increased exposure to GCP CBT among research staff (e.g., increased number of staff from a given study site completing CBT) may be correlated with increase in GCP use.
- **Focus on a specific type of clinical trial to examine GCP compliance.** Due to the varied nature of NIAID-funded clinical trial studies and multiple influences, focusing the examination of GCP use on a specific type of study (e.g., by Division, program, study goals) at a given time (e.g., new clinical trial startup, thus measuring GCP use at all points in trial conduct process) can facilitate a more thorough understanding and conclusions of GCP use increase and adherence.

Goal 8[10]: Publications in peer-reviewed journals on GCP CBT and GCP compliance (Long-term goal)

Performance Measure: Number of publications—manuscripts submitted or published—about GCP implementation and program evaluation at the end of the evaluation period.

Summary

NOVA conducted this evaluation feasibility study to determine whether conducting an evaluation of NIAID DCR's GCP CBT program is appropriate and to identify optimal evaluation approaches. In conjunction with the Evaluation Planning Group, the NOVA evaluation team engaged in a systematic assessment of evaluation options, methodologies, and instruments to provide guidelines and specific recommendations to design a process and outcome evaluation. After conducting the Feasibility Study, NOVA has come to the following conclusions:

a. A full-scale evaluation of the GCP CBT is feasible and should be conducted to determine the effectiveness of the training in increasing knowledge of GCP among target audiences.

[10] Goal 8 on publications is part of both Evaluation Components A: Evaluation of GCP CBT Program Effects and B: Examination of GCP Compliance.

b. Two distinct evaluation components can be assessed. They complement each other by providing a general picture of NIAID's focus on GCP use in the conduct of clinical trial research:

- **Evaluation Component A:** Evaluation of GCP CBT effects which will assess the influence of the training on knowledge acquisition of GCP
- **Evaluation Component B**: Examination of GCP compliance which will explore adherence to GCP in clinical trial research over time.

c. Data from each evaluation component will provide useful information.

KEY INFORMATION TO BE LEARNED FROM THE FULL-SCALE EVALUATION	
From **Evaluation Component A**: Evaluation of GCP CBT Effects	From **Evaluation Component B**: Examination of GCP Compliance
✓ Level of awareness of GCP CBT among target users ✓ Current knowledge of GCP among users (e.g., strengths and weaknesses) ✓ Overall reaction to GCP CBT (e.g., satisfaction) ✓ Perceived relevance of GCP CBT to users' work setting ✓ Effectiveness of GCP CBT in increasing knowledge of GCP among users ✓ Suggestions to improve effectiveness of GCP CBT	✓ GCP compliance in NIAID-funded clinical trial research studies over time ✓ Barriers, facilitators, and lessons learned in adherence to GCP guidelines ✓ Improvement in subject safety, quality of research data, and overall research efficiency over time, after availability of GCP CBT ✓ Suggestions to facilitate GCP compliance (identified by type of clinical trial study—e.g., R34s, and study stage—e.g., initiation)

All recommendations included in this report are summarized in **Appendix F**.

References

Centers for Disease Control and Prevention. (1999). Framework for program evaluation in public health. *MMWR, 48***,** RR-11.

Kirkpatrick, D. L. (1975). Techniques for evaluating programs. Parts 1, 2, 3 and 4. *Evaluating training programs*. ASTD.

Patton, M. Q. (1997). *Utilization-focused evaluation: the New Century Text* (3rd ed.). City, CA: Sage Publications.

APPENDIX A: NIAID GCP CBT Logic Model

Overall goal: The goal of the NIAID GCP CBT program is to increase knowledge about and use of GCP in planning, conduct, closeout, and reporting of clinical trials involving human subjects, demonstrating value added in subject safety, data integrity, and research efficiency.

CONTEXT	IMPLEMENTATION		OUTCOMES	
Resources/Inputs	Activities	Outputs	Short-term Outcomes	Long-term Outcomes
NIAID GCP CBT Steering Committee and Subcommittees • DCR staff • SPEB staff • Representatives from all NIAID Divisions • NCRS • EduNeering staff • Other NIAID contractors **NIAID funding**	**Curriculum Development** • Develop and refine learning objectives • Review content courses • Pilot test GCP CBT • Modify CBT based on findings from pilot test **GCP CBT Implementation** • Design and implement dissemination plan to inform target audiences (e.g., program officers, clinical investigators, pharmacists) of CBT • Implement GCP CBT	**GCP CBT Curriculum** • Completed GCP CBT curriculum **GCP CBT Dissemination** • Number of communication activities by type (e.g., e-mail announcements) and target audience (extramural/intramural researchers, grantees, contractors) • Number reached by type of audience **GCP CBT Usage** • Number of target users who access CBT Web site • Number of target users who access at least one course • Number of target users who complete courses • Courses most frequently accessed, started, and completed	**<u>Evaluation A: GCP CBT Effects</u>** **Awareness/ Usage** • Target audiences are aware of the GCP CBT course • Increase use of GCP CBT by target audiences **Reaction to Training** • Users are satisfied with CBT • Users perceive CBT as relevant and transferable to their work setting **Learning** • Users demonstrate increase GCP knowledge	**<u>Evaluation B: GCP Use</u>** **Transfer of Knowledge** • Increased application of GCP in the conduct of clinical trial research **Results of GCP Use** • Improved subject safety • Trustworthy research data • Increased research efficiency (i.e., bringing a research project from the research question to the answer expeditiously, improved communication) • Publications on GCP CBT and GCP compliance

Contextual/environmental factors (e.g., NCRS consideration of mandatory GCP CBT, policy, monitoring practices, other types of GCP training)

APPENDIX B: NIAID GCP CBT Evaluation Matrix

Overall goal: The goal of NIAID's GCP CBT program is to increase knowledge and use of GCP in planning, conduct, closeout, and reporting of clinical trials involving human subjects, and demonstrate value added in three areas: subject safety, data integrity, and research efficiency.

Evaluation Component A: **Evaluation of GCP CBT Effects**

Evaluation Questions Addressed	Information Required	Information Source(s)	Data Collection Methods	Data Analysis Methods	Limitations	What the Analysis will Allow to Say
GCP CBT Awareness and Usage						
Are target audiences aware of the availability of and way to access the GCP CBT course?	# and type of GCP CBT dissemination activities (including to whom disseminated) % of known target audience aware of GCP CBT # of target audience who access/open CBT Web site # target audience who access at least one CBT topic	NIAID GCP CBT team NIAID alerts on regulations about clinical trial research Target audience EduNeering data	Tracking sheet on dissemination activities from NIAID GCP CBT team Periodic (e.g., annual) cross-sectional surveys User Profile data CBT use log	Descriptive analysis	Unknown total N for target audience	Impact of dissemination activities Appropriateness of plan (e.g., type of activities, messages, intensity) to affect awareness of or access to CBT Participation in CBT
What is the rate of completion of GCP CBT courses among target audiences?	# topics started per trainee # topics completed per trainee # intramural researchers # extramural researchers # NIAID staff (other than intramural researchers)	EduNeering data Target audience	User Profile data CBT use log	Descriptive analysis		Participation in GCP CBT among target audience

Evaluation Questions Addressed	Information Required	Information Source(s)	Data Collection Methods	Data Analysis Methods	Limitations	What the Analysis will Allow to Say
Reaction to Training						
What is the overall reaction of trainees to GCP CBT?	Satisfaction with GCP CBT Participation in GCP CBT Comments in help desk	CBT trainees EduNeering data	User Satisfaction Survey CBT use log Help desk available on EduNeering Web site Focus groups	Quantitative and qualitative descriptive analyses	Self-reported data Survey and comments are optional measures.	Overall satisfaction about GCP CBT CBT areas for improvement
How relevant and transferable is GCP CBT to trainees' work setting?	Perceived relevance, usefulness of GCP (e.g., facilitates their work or research) to trainee's job responsibilities and tasks Comments in help desk	CBT trainees	User Satisfaction Survey Help desk available on EduNeering Web site Focus groups	Quantitative and qualitative descriptive analyses	Self-reported data Optional measures	Perceived applicability of CBT content to NIAID clinical trial research CBT areas for improvement
Learning						
What is the influence of GCP CBT training on trainees' knowledge related to GCP?	Knowledge of GCP Identification of GCP components Items on application of GCP to real-life scenarios	CBT trainees Comparison groups[11]	Pre- and post-tests • CBT trainees • Comparison groups	Descriptive, comparative, and correlational analyses	Self-reported data	GCP knowledge among target audiences GCP knowledge associated with CBT

1 Comparison groups are members of the target audience who (a) failed the pretest and did not complete CBT and (b) did not take the pretest or any CBT.

Evaluation Component B: Examination of GCP Compliance

Evaluation Questions Addressed	Information Required	Information Source(s)	Data Collection Methods	Data Analysis Methods	Limitations	What the Analysis will Allow to Say
Transfer of GCP Knowledge						
Is there an increase in the application of GCP over time after the availability of GCP CBT?	Perceptions of site monitors, NIAID project oversight staff, and grantee staff on GCP compliance Data on protocols and procedures that reflect compliance with GCP	Site monitors, NIAID project oversight staff (e.g., project managers), and research staff	Site Monitoring Tool Survey on issues related to GCP use	Descriptive, comparative, and correlational analyses Qualitative analysis	Data quality and reliability (external factors influencing GCP compliance)	Trend in GCP compliance when conducting clinical trial research over time
What are common barriers and facilitators of GCP compliance?	Perceptions of site monitors, NIAID project oversight staff, and grantee staff on GCP compliance, barriers, and facilitators	Site monitors, NIAID project oversight staff (e.g., project managers), and research staff	Site Monitoring Tool Survey on issues related to GCP use Focus groups	Quantitative/ qualitative descriptive analyses	Data quality and reliability Self-reported data	Most common barriers encountered to implement GCP Courses of action that facilitate GCP compliance
Results of GCP Compliance						
Is there an improvement in subject safety over time after availability of CBT?	Data on protocols and procedures that reflect subject safety GCP guidelines	Site monitors	Site Monitoring Tool	Descriptive statistics	Data quality, reliability	Changes in subject safety over time
Is there an improvement in the quality of research data over time after availability of CBT?	Protocols, procedures, and reports that reflect use of GCP in the collection and management of data	Site monitors	Site Monitoring Tool	Descriptive statistics	Data quality, reliability	Changes on data quality over time
Is there an improvement in research efficiency over time after availability of CBT?	Perceptions regarding influence of GCP to facilitate conduct of research (e.g., easy to follow guidelines, improved communication)	NIAID Division representatives NIAID Program officials, Site monitors, grantees	Interviews with NIAID Division representatives Focus groups Site monitoring data over evaluation period	Descriptive statistics Qualitative analysis	Data quality and reliability	Overall perceptions of research efficiency over time

APPENDIX C: Proposed Evaluation Timeline

The following is the timeline for the evaluation work plan which specifies key evaluation activities and the timeframe to complete them.

The proposed timeline will list key evaluation activities. Activities that pertain to each evaluation component will be noted.

Evaluation Component A: Evaluation of GCP CBT Effects

Evaluation Component B: Assessment of GCP Compliance

	Evaluation Component A	Evaluation Component B
Year 1		
Focus of evaluation: Intramural research, domestic clinical trials		
Form and convene an Evaluation Advisory Committee.	✓	✓
Update the program logic model.	✓	✓
Finalize evaluation study questions.	✓	✓
Identify benchmarks and performance indicators.	✓	
Update the evaluation planning matrix.	✓	✓
Conduct a pilot study on CBT and evaluation measures/procedures.	✓	
Conduct content analysis of focus groups.	✓	
Report to program staff and Evaluation Advisory Committee on focus group findings, conclusions, and recommendations.	✓	✓
Incorporate suggestions provided during focus groups regarding evaluation design and procedures.	✓	✓
Finalize measures of GCP compliance.		✓
Collect baseline data on GCP compliance (e.g., Site Monitoring Tool, survey on barriers to GCP compliance).		✓
Collect pre- and post-test data on GCP knowledge.	✓	
Collect process data on GCP CBT program implementation and operations.	✓	
Conduct regular meetings with the NIAID GCP CBT program workgroup.	✓	✓

	Evaluation Component A	Evaluation Component B
Conduct quarterly or semi-annual meetings with the Evaluation Advisory Committee.	✓	✓
Produce an annual report on program operations and interim findings (e.g., level of CBT participation).	✓	✓
Year 2		
Focus of evaluation: Intramural research, domestic clinical trials		
Analyze and report to program staff and Evaluation Advisory Committee on knowledge outcome data and initial findings from baseline data on GCP use.	✓	✓
Continue collection of process data on program operations.	✓	
Continue collection of pre- and post-test data on GCP knowledge.	✓	
Update the program logic model.	✓	
Conduct regular meetings with the NIAID GCP CBT program workgroup.	✓	✓
Conduct quarterly or semi-annual meetings with the Evaluation Advisory Committee.	✓	✓
Provide support for dissemination activities on the GCP CBT Program (e.g., program presentations).	✓	
Produce an annual report on program operations, interim findings, and initial outcomes (e.g., GCP knowledge, GCP compliance).	✓	✓
Year 3		
Focus of evaluation: Extramural research, domestic clinical trials		
Update the evaluation planning matrix.	✓	✓
Conduct focus groups with project oversight and research staff on barriers to GCP compliance and CBT.	✓	✓
Conduct interviews with NIAID Division representatives, site monitors, and site PIs on the applicability of GCP to clinical trial research projects.		✓
Collect follow-up data on GCP compliance (e.g., Site Monitoring Tool, survey on barriers to GCP compliance).		✓
Collect pre- and post-test data on GCP knowledge.	✓	

	Evaluation Component A	B
Collect process data on program operations.	✓	
Analyze and report to program staff and Evaluation Advisory Committee on program process, knowledge outcome data, and initial findings from baseline data on GCP use.	✓	✓
Conduct regular meetings with the NIAID GCP CBT program workgroup.	✓	✓
Conduct quarterly or semi-annual meetings with the Evaluation Advisory Committee.	✓	✓
Provide support for dissemination activities on the GCP CBT program (e.g., program presentations).	✓	✓
Produce an annual report on program operations, interim findings, and initial outcomes (e.g., GCP knowledge, GCP compliance).	✓	✓
Year 4		
Focus of evaluation: International clinical trials		
Collect follow-up data on GCP compliance (e.g., Site Monitoring Tool, survey on barriers to GCP compliance).		✓
Collect pre- and post-test data on GCP knowledge.	✓	
Collect process data on program operations.	✓	
Analyze and report to program staff and Evaluation Advisory Committee on program process, knowledge outcome data, and findings on GCP use.	✓	✓
Regular meetings with NIAID GCP CBT program workgroup.	✓	✓
Regular meetings with Evaluation Advisory Committee.	✓	✓
Provide support for writing manuscripts and dissemination activities on the GCP CBT program (e.g., program presentations).	✓	✓
Produce an annual report on program operations, interim findings, and initial outcomes (e.g., GCP knowledge, GCP compliance).	✓	✓
Year 5		
Conduct interviews with NIAID Division representatives.		✓
Conduct focus groups with project oversight and research staff on barriers to GCP		✓

	Evaluation Component A	Evaluation Component B
compliance and CBT.		
Conduct content analysis of qualitative data gathered through the interviews and focus groups.	✓	✓
Provide support for writing manuscripts and dissemination activities on the GCP CBT Program (e.g., program presentations).	✓	✓
Finalize analysis and report to program staff and Evaluation Advisory Committee on program process, knowledge outcome data, and findings on GCP use over the years.	✓	✓

APPENDIX D: NIAID User Profile Questionnaire

1. Privacy statement informing the user of what information will be maintained, secured, and reported on. (Language to be drafted/edited at a later time). User will be given two choices.
   - ❒ Agree – continue user profile and onto EduNeering
   - ❒ Disagree – exit user profile

2. Which of the following best describes your relationship with NIAID?
   - ❒ NIAD Employee
   - ❒ NIAID contractor or staff or subcontractor of NIAID contractor
   - ❒ NIAID grantee or staff or subcontractor of NIAID grantee
   - ❒ NIAID collaborator or staff or subcontractor of NIAID collaborator (non-grant/contract collaboration)

3. With which NIAID Division(s) do you work or collaborate (grantees and collaborators, choose all that apply)?
   - ❒ Division of Acquired Immunodeficiency Syndrome (DAIDS)
   - ❒ Division of Allergy, Immunology, and Transplantation (DAIT)
   - ❒ Division of Clinical Research (DCR)
   - ❒ Division of Extramural Activities (DEA)
   - ❒ Division of Intramural Research (DIR)
   - ❒ Division of Microbiology and Infectious Diseases (DMID)
   - ❒ Office of the Director (OD)
   - ❒ Office of Management and Operations (OMO)
   - ❒ Vaccine Research Center (VRC)

4. Which of the following describe your role(s) in NIAID-sponsored clinical trials (<u>check all that apply</u>)?
   - ❒ Principal Investigator
   - ❒ Investigator (Associate Investigator, Assistant Investigator, Co-Investigator, Sub-Investigator)
   - ❒ Physician
   - ❒ Nurse
   - ❒ Clinical Research Coordinator, Study Coordinator, Site Coordinator, Project Manager
   - ❒ Data Manager
   - ❒ Technician (lab, radiology, ultrasound, etc.)
   - ❒ Pharmacist, Pharmacy Assistant
   - ❒ Program Officer, Program Specialist, Grant Manager
   - ❒ Regulatory Affairs Manager/Staff
   - ❒ Clinical Research Associate
   - ❒ Protocol Specialist
   - ❒ Safety Monitor, Safety Specialist, Medical Monitor
   - ❒ Research Assistant
   - ❒ Training Manager/Staff
   - ❒ Biostatistician
   - ❒ Administrative Staff (clerk, receptionist)
   - ❒ Member of an Oversight Board (i.e., Ethics Committee, Data Safety Monitoring Board, Safety Monitoring Committee, Endpoint Review Committee)

5. Please choose the phrase that **<u>best describes</u>** your current career stage.
   - ❒ Undergraduate student or intern
   - ❒ Graduate or post-baccalaureate student
   - ❒ Postdoctoral, fellow or intern
   - ❒ Clinical practitioner—certified/licensed
   - ❒ Professional/nonacademic
   - ❒ Faculty—instructor, assistant or associate professor
   - ❒ Faculty—full professor, chair, dean, etc.
   - ❒ Other

6. Approximately how many hours of formal Good Clinical Practice (GCP) or Human Subjects Protection (HSP) training (online or classroom) have you completed in the past?
   - ❒ <2 hours
   - ❒ 2-10 hours
   - ❒ >10 hours

7. Approximately how many years have you been involved in the conduct of clinical research?
   - ❒ <2 years
   - ❒ 2-8 years
   - ❒ >8 years

8. Please select any of the following degrees or certificates that you currently hold (check ALL that apply).
   - ❒ None
   - ❒ ACRP certification for Clinical Research Associate (CRA)
   - ❒ ACRP certification for Clinical Research Coordinator (CRC)
   - ❒ ACRP certification for Clinical Trial Investigator (CTI)
   - ❒ APPI certification for Certified Physician Investigator (CPI)
   - ❒ SOCRA certification for Clinical Research Professionals (CCRP).
   - ❒ RAPS Regulatory Affairs Certification (RAC)
   - ❒ A university degree (associate, bachelor's, master's or PhD) in Clinical Research
   - ❒ Certificate of Attendance for a computer-based GCP course within the past year
   - ❒ Certificate of Attendance for a classroom GCP course within the past year
   - ❒ Other

9. Please choose the sentences that best describe your current knowledge of GCP.
   - ❒ My knowledge is excellent. My colleagues use me as a resource.
   - ❒ My knowledge is good. I can conduct my job duties according to GCP independently.
   - ❒ My knowledge is fair. I regularly consult with others.
   - ❒ My knowledge is poor. I don't understand what GCP is.

10. In what geographic region(s) are you, personally, involved in clinical research (check ALL that apply)?
   - ❒ Asia and the Pacific
   - ❒ Europe and Central Asia
   - ❒ Latin America and the Caribbean
   - ❒ Middle East and North Africa
   - ❒ North America

APPENDIX E: GCP CBT User Satisfaction Survey

Please rate the following:

Question 1: *(transfer of knowledge/relevance)*	Very Irrelevant	Somewhat Irrelevant	Somewhat Relevant	Very Relevant
How relevant was the training to your job responsibilities?	1	2	3	4
Question 2: *(learning/knowledge)*	Not Informative at All	Not very Informative	Somewhat Informative	Very Informative
How informative was the training in allowing you to understand NIAID GCP guidelines (e.g., improved knowledge, clarified misconceptions)?	1	2	3	4
Question 3: *(reaction/satisfaction with system design)*	Very Difficult	Somewhat Difficult	Somewhat Easy	Very Easy
How would you rate the ease of navigation of the online training courses?	1	2	3	4
Question 4: *(reaction/satisfaction with system design)*	Very Dissatisfied	Somewhat Dissatisfied	Somewhat Satisfied	Very Satisfied
How satisfied were you with the training delivery (i.e., presentation of material on computer screen)?	1	2	3	4
Question 5: *(reaction/satisfaction with content)*	Very Dissatisfied	Somewhat Dissatisfied	Somewhat Satisfied	Very Satisfied
How satisfied were you with the GCP CBT training content?	1	2	3	4
Question 6: *(reaction/overall satisfaction)*	Very Dissatisfied	Somewhat Dissatisfied	Somewhat Satisfied	Very Satisfied
In general, how satisfied were you with the GCP CBT training (i.e., the course was engaging, easy to understand, on target)?	1	2	3	4

APPENDIX F: Summary of Recommendations for a Comprehensive Evaluation of the GCP CBT Program that Includes Evaluation Component A: *Evaluation of GCP CBT Effects* and Evaluation Component B: *Examination of GCP Compliance*

Recommended Evaluation Steps
Step 1 – Engage Stakeholders ➢ Involve appropriate stakeholders in the evaluation. Key stakeholders include NIAID leadership and staff within DCR and SPEB, members of the Evaluation Advisory Committee, other NIAID Divisions in charge of clinical trials, and the clinical trial research community.
Step 2 – Describe the Program ➢ Review the program logic model annually to accurately reflect program changes.
Step 3 – Focus the Evaluation Evaluation Advisory Committee: ➢ Committee members should have experience in evaluation (i.e., evaluation of GCP, HSP, e-learning, adult learning strategies), GCP, NIH-related clinical trial regulations, and clinical trial monitoring. Evaluation questions: <u>For Evaluation A: Evaluation of GCP CBT Effects</u> ➢ Evaluation questions can address program awareness and usage, overall reaction to training, and learning of GCP. <u>For Evaluation B: Examination of GCP Compliance</u> ➢ Evaluation questions can explore GCP adherence over time, related barriers and facilitators, and improvement in subject safety, quality of data, and overall research efficiency. Benchmarks and performance indicators for GCP CBT: ➢ Establish realistic benchmarks for training process and outcome indicators. Timeline: ➢ Allow enough time for a full-scale evaluation. A 5-year plan is an appropriate time period to assess both evaluation components. Plan for incremental steps in its scope (e.g., focus on intramural research the first 2 years and then assess extramural research; examine domestic clinical trials first and then international studies; and collect longitudinal data over a reasonable period to examine GCP use). Methods: ➢ Use a mixed-method evaluation design of qualitative and quantitative techniques to obtain a full picture of GCP CBT effects and GCP compliance. ➢ Use power analysis (for effect size) that specifies sample size to determine statistically significant changes in GCP CBT effects.

- Minimize respondent burden and ensure confidentiality or anonymity of sensitive personal information.

Step 4 – Gather Credible Evidence

Data Sources:

- Data collection instruments should have the input of the Evaluation Advisory Committee and pertinent audiences (e.g., site monitors, program staff).

  For Evaluation A: Evaluation of GCP CBT Effects, collect:

  - Data on participants' characteristics (e.g., role at NIAID, previous GCP experience)
  - Item-level responses by user ID for both pre- and post-test in order to better assess change in knowledge
  - Descriptive metrics on GCP test-retake: number of people retesting, number of times retesting, test scores at each test taking (for course, module, and "challenge" post-test)
  - CBT participation data (number of users who access/open CBT Web site; number of users who access at least one course of GCP CBT; courses most frequently accessed, started, and completed by users).

  For Evaluation B: Examination of GCP Compliance, collect:

  - Data on key GCP data elements, perceptions regarding barriers and facilitators of GCP use, subject safety, data quality, and overall research efficiency.
  - Related data from site monitors, study staff, and NIAID representatives, where appropriate.

Answering Evaluation Questions on Evaluation Component A: Evaluation of GCP CBT Effects

Goal 1: Target audiences are aware of GCP CBT.

- The program dissemination plan should be very specific, describing number of expected activities directed to whom, by whom, and by when, and identifying benchmarks for activities (e.g., minimum number of activities by type planned to implement).

Goal 2: GCP CBT users have favorable reactions to the training.

- Conduct a pilot of GCP CBT including administration of evaluation instruments and focus groups.
- Set benchmarks for expected ratings of overall satisfaction with CBT. Set criteria should specify (a) minimum rating score that would demonstrate CBT satisfaction and (b) expected proportion of users expected to score at least the minimum satisfaction score.
- Survey items should collect information that will be of use to NIAID program staff and stakeholders.

Goal 3: GCP CBT increases knowledge of GCP.

- Select an evaluation design that minimizes threats to validity of evaluation findings (cons).
- Consider an evaluation design that includes a comparison group with pre- and post-tests, as it can best attribute differences in GCP knowledge to CBT.
- If a comparison group is not possible, ensure mandatory pretest and post-test (at least temporarily for the pretest until targeted sample size for statistically significant effects is reached).
- Determine expected effect size on knowledge change (realistic) to establish appropriate sample size.

➢ Collect data at an individual level. Matching a person's pretest score to the person's post-test score allows for a more precise analysis of change in knowledge.
Answering Evaluation Questions on Evaluation Component B: Examination of GCP Compliance
Goal 4: Increased application of GCP in conduct of clinical trial research Goal 5: Improved subject safety Goal 6: Improved quality and integrity of clinical research data Goal 7: Improved research efficiency ➢ Collect data on GCP compliance at least three times (e.g., baseline in Year 1, at the end of Year 3, and at the end of Year 4) to better assess any change in GCP use. ➢ Collect baseline data before GCP CBT is implemented. ➢ In the case of "late baseline" data, collect data soon after program startup. ➢ Collect GCP compliance data on new clinical trial studies only to better assess any potential influence of the program on GCP use.

APPENDIX G: NIAID GCP CBT Review of Literature

Purpose

To assess published literature on training related to GCP and human research protection. The search included literature on GCP training and evaluations of GCP computer-based training and was expanded to include computer-based training relating to research conduct more generally,; focused on documented changes in behavior.

Key Search Terms

Computer based* training +Results/Outcomes/Impact	Quality Assurance + clinical trials
Good clinical practice/GCP + Results/Outcome/Impact	NIH + computer based training
Good clinical practice/GCP + monitoring	VA + Site Monitoring and Review Team (SMART)
Good clinical practice/GCP + training	VA + Good Clinical Practices Monitoring Group (GCPMG)
Good clinical practice/GCP + Computer based education	E-learning
Good clinical practice/GCP + adherence	Web-based training in research
Good clinical practices/GCP + education	Ethics training
Good clinical practice/GCP + FDA + regulations	Web-based ethics training
Internal Conference Harmonization (ICH)	
Human research protection + training + Results/Outcome /Impact	

Sources Used for Literature Review

Pub Med	Department of Health and Human Services (DHHS) Web site
Veterans Affairs (VA)	Google and Google Scholar
National Institutes of Health (NIH)	Association of Clinical Research Professionals (ACRP)
Food and Drug Administration (FDA)	Ancestor search on all articles
American Psychological Association (APA)	

Summary of Results

467 abstracts were identified and reviewed
50 abstracts were related to broad category of computer-based training, but were not evaluation studies or were not related to human subject protection or GCP
15 articles were reviewed
6 papers met literature search requirements (related to GCP, evaluation studies, or Web-based training)
3 papers reviewed the literature and current practice related to GCP and/or human subjects protection
2 papers reviewed clinical research practice
1 paper evaluated a Web-based program on responsible conduct of research

Searched and Contacted Organizations that Do Computer-Based Training on GCP and/or HSP

Department of Health and Human Services (DHHS)	Several universities that use the CITI (Collaborative IRB Training Initiative) training
National Institutes of Health (NIH)	Co-founder of the CITI program
Office of Human Subject Research Protection (OHRP)	Association of Clinical Research Professionals (ACRP)
Office of Research Services (ORS) Division of Radiation Safety (DRS)	

Review of Documents

Type of Document: Review Papers
Citation: Wood A, Grady C, Ezekiel JE. The crisis in human participants research: identifying the problems and proposing solutions. 2002 September. http://www.bioethics.gov/backround/emanuelpaper.html (Presented at the Presidents Council on Bioethics)
This article describes in detail the problems faced in the process of protecting human subjects for research and proposes several solutions that would comprehensively address these problems. The authors describe problems in three major areas that are inherent in the protection of human subjects in research: 1. Problems with the structure of the human resource participants protection system 2. Review process 3. Problems in performance assessment NIH initiated reforms such as accreditation, credentialing of IRB personnel, legislative actions, OHRP conflict of interest, quality initiatives, and review of central IRBs. While NIH has made efforts to address these problems and issues, they fail to address the structural problems affecting research review.
Citation: Emanuel EJ, Wood A, Fleischman A, Bowen A, Getz KA, Grady C, Levine C, Hammerschmidt DE, Faden R, Eckenwiler L, Muse CT, Sugarman J. Oversight of human participants research: identifying problems to evaluate reform proposals. Ann Intern Med. 2004 Aug 17; 141(4):282-91.
This paper builds on the Wood et al., 2002 article. The authors describe in detail the problems and deficiencies in the oversight of research involving human participants. Six fundamental components of a more effective reform proposal are evaluated in the 12 problem areas: 1. Accreditation 2. Credentialing of IR professionals 3. Central IRBs 4. Legislative proposals 5. OHRP initiatives 6. IOM Report

Citation: Rosenbaum JR. Educating researchers: ethics and the protection of human research participants. Critical Care Med. 2003 Mar 31;(3 suppl):S161-6.

In this review paper, the author discusses research ethics education and training of researchers in human subjects research. The author poses many good questions about how well ethics training creates changes in behavior. Her conclusions include the following:

- Training programs should have clearly articulated goals.
- Training success should be objectively measured.
- Educators should be clear about whether they seek to improve or assess knowledge, confidence, skills, attitudes, or other determinants of behavior.
- It has been difficult to demonstrate that educational interventions that improve knowledge also have a positive impact on actual behavior, despite evidence from the psychological literature of their relationship.
- Having a formal curriculum is only one of the many complex influences on behavior.
- Quizzes should measure change in performance. How does this happen?
- Better assessments will cost more.

Citation: Califf RM, Morse MA, Wittes J, Goodman SN, Nelson DK, DeMets DL, Iafrate RP, Sugarman J. Toward protecting the safety of participants in clinical trials. Control Clin Trials. 2003 Jun;24(3):256-71.

The authors review limitations of clinical research oversight to ensure the safety of participants in clinical trials at major institutions and offer recommendations for improving the current system.

Based on their findings, the authors recommend that all research staff should participate in formal training that includes GCP, basic ethics and policies, and practical aspects of clinical trial management, including record keeping and documentation. Simple certification of completion of a training course would increase training participation.

Type of Document: Training and Evaluation Papers
Citation: Sather MR, Raisch DW, Haakenson CM, Buckelew JM, Feussner JR. Promoting good clinical practices in the conduct of clinical trials: experiences in the Department of Veterans Affairs Cooperative Studies Program. Control Clin Trials. 2003 Oct;24(5):570-84.
This research study evaluated changes in adherence to good clinical practice (GCP) by the VA's Cooperative Studies Program, as measured by results of formal site visit reviews conducted over a 3-year period by a GCP Review Group. A trained reviewer conducted GCP site reviews. Data were compared between the 2-year implementation period and continuing follow-up period. Evaluation measures included: • Assessment tool – 62 ICH-derived GCP key elements related to eight GCP focus areas (e.g., patient consent issues, safety monitoring, IRB); each question had three possible responses: pass, fail, or not applicable • Overall summary scale – compilation of assessment of 14 GCP-selected items identified as critical (e.g., consent obtained prior to study procedures, annual IRB review obtained without lapse) • Overall GCP performance – reviewers provided summary evaluation of overall GCP performance of each site, divided into three categories: high, average to good, below average
Citation: You YN, Jacobs L, Martinez ED, Budinger SC, Wittlief EJ, Myles SK, Ota DM, American College of Surgeons Oncology Group. Improved surgeon performance in clinical trials: an analysis of quality assurance audits from the American College of Surgeons Oncology Group. J Am Coll Surg. 2006 Sep;203(3):269-76.

To assess the quality of trial conduct by surgeons on a national level and the feasibility of improvement through education, this study examined the findings of the Quality Assurance Audit Program of the American College of Surgeons Oncology Group (ACOSOG) over time. ACOSOG is a NCI-sponsored national oncology cooperative group with the mission to evaluate surgical management of malignant solid tumors in a multidisciplinary setting.

Before participation in ACOSOG, many surgeons were unfamiliar with the theory and practice of clinical trials. To assess the quality of conduct of surgical oncology trials nationwide and benchmark progress over time, the experience of the ACOSOG quality assurance audit program was reviewed. The purpose was to determine whether trial performance in regulatory requirements and patient case review improved in response to quality assurance audits and other educational programs.

The audit team reviewed and verified submitted data against source documents, including original IRB documents, informed consents, NCI Drug Accountability Record Forms, medical records (clinical notes, diagnostic test reports, laboratory data, procedural reports, and so on), research records signed and dated by study personnel, and subject diaries or calendars.

Deviations found were rated as either a "major deficiency" (defined as a variance from protocol-specific procedures that made the resulting data questionable) or a "minor (or lesser) deficiency" (defined as one that was judged not to have a significant impact on the outcomes or interpretation of the study). Additionally, each audit component (REG, PHARM, or PCR) was given an overall grade of "acceptable," "acceptable needing follow-up," or "unacceptable."

Outcomes of 249 routine audits conducted from 2001 to 2004 were reviewed for major and minor deficiencies and overall performance (acceptable versus unacceptable) in compliance with regulatory requirements (REG) and patient case review (PCR).

Conclusions:

In response to educational programs, surgeons' performance in clinical trials has measurably improved.

Findings of the study highlight an association between educational programs and improvements in trial performance over time.

Citation: Barnes BE, Friedman CP, Rosenberg JL, Russell J, Beedle A, Levine AS. Creating an infrastructure for training in the responsible conduct of research: The University of Pittsburgh's experience. Acad Med. 2006 Feb;81(2):119-27.

This paper describes the University of Pittsburgh's experience in the design, implementation, and evaluation of a Web-based, institution-wide training program on responsible conduct of research.

Evaluation measures included:

- Completion rate for each training module
- Users' satisfaction with content and ease of use of program
- Users' questionnaire responses assessing perceptions of quality, educational effectiveness, and appropriateness of content
- Users' performance on module tests
- Log files of users' activities (e.g., module(s) accessed)
- Number of module certifications issued.

Review of Literature: Related Programs on GCP CBT or Human Subjects CBT

NIH COMPUTER BASED TRAINING: Protecting Human Subjects

Contacted: Benita Bazemore, Program Specialist, National Institutes of Health, Clinical Center, Office of the Director

Allison Wichman, Director NIH Intramural Human Subject Research, National Institutes of Health, Office of Human Subject Research, and author of the course.

Evaluation of CBT: No published studies; no evaluation data. Anecdotally, they feel that knowledge has increased and they can tell by the calls that come into the office that people are better informed.

Collaborative Institutional Training Initiative (CITI)

Purpose: Web-based CBT on human subjects protection developed by the University of Miami and the Fred Hutchinson Cancer Research Center (https://www.citiprogram.org/citidocuments/aboutus.htm). Reportedly used in 400 institutions.

Contact: Paul G. Braunschweiger, Ph.D., Professor of Radiation Oncology, Director Office of Research Education, CITI Co-Founder. (Note: No response to date from Dr. Braunschweiger.)

Evaluation of CBT: No published studies.

Web-based Instruction on Informed Consent

Purpose: This CBT on informed consent is required for faculty and students conducting human research at the University of Minnesota. (http://www.research.umn.edu/consent/).

Contact: (Note: No response to date.)

Evaluation of CBT: No published studies.

ACRP (Association of Clinical Research Professionals) (http://www.acrpnet.org/)

Contact: Lisa Edy, Customer Service, ACRP

Purpose: Founded to address educational and networking needs of research nurses and others who support the work of clinical investigations. CRP offers continuing medical education that covers GCP topics and human subjects protection issues.

Certification: To be eligible for certification, an individual must be a research coordinator for 2 years. Certification involves completing an application and taking a 3-hour exam. Recertification is required every 2 years. Several types of certification are offered:

- CRA- Clinical Research Associate
- CRC- Clinical Research Coordinator (FDA)
- CTI- Clinical Trials Investigator
- CPI- Certified Physician Investigator.

Exams: In 2000, the ICH CTI Exam Committee was formed in order to develop an exam for clinical trial investigators in the ICH regions. The exam was first held in December 2001. A similar CTI Exam Committee has been established in North America. Their first CTI exam was held in April of 2002.

Exam reliability measurement: After every certification exam, the Academy's CRA Exam Committee reviews the results. Each item is analyzed for appropriate psychometric characteristics. Items with poor statistical results are reviewed by the Exam Committee to ensure that they have been scored properly. Participant feedback regarding the exam and its contents is also reviewed and taken into consideration when reviewing the exam and future test items.

Evaluation: No evaluation studies to date.

Recommendations for Project Implementation Based on Literature Search

Training

- Ensure that program content addresses desired learning objectives and NIH requirements.
- Make training easily accessible.
- Ensure that training is systematically deployed and delivered.
- Ensure that the program outcomes are sustainable.
- Define methods for updating and editing the course.

Communication

- Be able to provide results or data to relevant Institutes.
- Communicate to researchers that the course is relevant and valuable.
- Ensure that all partners, contacts, and stakeholders are engaged.

Data

- Develop ways of tracking completion.
- Create certification or completion databases.
- Ensure that the progress of the project can be accurately tracked.
- Ensure that user demographics are systematically collected.
- Determine whether a control or comparison group is needed.
- Consider how the stage of the program (newly developed program) will affect the evaluation.
- Ensure that indicators are relevant, understandable, measurable, and useful.

Recommendations for Evaluation Based on Literature Search

Indicators

- Clearly define measurable outcomes.
- Creatively define ways to measure intangibles like changes in attitude.
- Define ways to address questions like: Does CBT affect performance of those who take it and does it improve the protection of human subjects in research?
- Ensure that quizzes and satisfaction surveys accurately assess performance and satisfaction.
- Assess how implementation of the CBT will impact the evaluation.

Data

- Review satisfaction surveys from other studies like the Pittsburgh study.
- Clearly define methods for data collection.
- Identify how often data will be collected.
- Define who is responsible for data collection.
- Define how data will be managed, stored, and analyzed.

Analysis and Interpretation

- Define roles and responsibilities of key people in the analysis and interpretation of data.
- Define methods for analysis.

Communication of Results

- Identify and define the audience (who will be involved in drawing, interpreting, and justifying conclusions).
- Identify media for communicating information.
- Identify and define plans for using evaluation findings and how, when, and where findings will be used.

APPENDIX H: Summary of Interviews with Site Monitors

Purpose of Interviews. The NOVA evaluation team conducted interviews with site monitoring staff from NIAID Divisions sponsoring clinical trial research studies to identify existing data sources that might be used to evaluate the effects of NIAID's GCP CBT on the conduct of clinical trials. Specific attention was placed on identifying metrics of changes or improvements in the use of GCP that could potentially be tied to the GCP CBT training.

Responses to an interview protocol developed by NOVA with input from the Evaluation Planning Group were obtained from seven individuals representing various positions, titles, and Divisions within NIAID, including a Health Specialist, a Project Officer for a site monitoring contract, a Clinical Monitoring Coordinator, a Clinical Trials Manager, a Site Monitoring Contractor, and Clinical Research Oversight Managers. They represented NIAID's Division of AIDS; Regulatory Compliance and Human Subjects Protection Branch; Office of Clinical Research Affairs, Division of Microbiology and Infectious Diseases, and private companies that provide clinical trials site monitoring[12].

The interview protocol queried about site monitoring practices for GCP compliance, with the intent to identify "core" or common key elements that are regularly monitored and collected across Divisions. The site monitoring interviews took place from March 2007 through May 2007. A summary of these findings is described below.

Site Monitoring Process

The monitoring process is detailed, intricate, and complex, involving various individuals (e.g., ranging from research staff, independent monitoring contractors, and NIAD monitoring staff) with different roles, responsibilities, and levels of participation at different stages in the monitoring process. Moreover, monitoring practices tend to vary based on NIAID's Divisions, programs, and clinical trial studies or whether the research is intramural or extramural. For example, different systems are employed monitoring intramural and extramural research; monitoring of the former tends to be more detailed than the latter. An example of monitoring steps followed by staff at one Division are summarized:

1. A prestudy and/or site assessment visit usually occurs before the initial site monitoring visit. Assessment visits are conducted to examine basic adequacy of infrastructure for a study.
2. The Clinical Monitor Coordinator identifies the risk level, complexity, and characteristics of the study protocol.
3. The monitoring plan and templates are developed by the monitor with the guidance of the Coordinator and input from a Scientific Champion (they are in direct and regular contact with the study PI).
4. Site monitoring takes place and a monitoring report is produced.
5. The report is reviewed by the Coordinator, who prepares an annotated version and communicates with the Scientific Champion about report findings.

[12] Telephone interviews were conducted with six individuals; written responses to the interview guide were obtained from one.

6. Recommendations are made by the Coordinator and integrated into the site monitoring plan. The Scientific Champion is in regular communication with the study PI about issues and purported solutions.

- Monitoring is done with attention to compliance with GCP and applicable regulatory requirements, including ICH GCP guidance for monitoring as noted in section 5.18.4 as well as 21 CFR part 312.56 and 312.57. DHHS regulations at 45 CFR part 46 are also observed for the protection of human subjects (e.g., informed consent procedures, IRB compliance, and additional protections for vulnerable populations).
- Monitors look at all the labs and documents. Sometimes study subjects do not report properly, creating the impression that the PI did something incorrectly. The monitor's role is to determine who is right and work with the study coordinator and the PI to clarify discrepancies.
- One of the monitoring functions is verification of source data (e.g., lab reports, regulatory documents, pharmacy and laboratory documents, physical exams, medical records, and discharge summaries) in which these data are evaluated against study-specific protocol requirements and GCP standards.
- Accuracy and completeness of trial data are also verified through Case Report Forms (CRFs). CRFs are used to enter study data, which are then compared with source documents. During a site visit, monitors re-enter data and a report is generated that checks for discrepancies between data entered by research staff and data entered by monitors.
- The monitoring plan specifies what will be examined and compared or contrasted (e.g., subjects, sections, information), which varies according to study characteristics and implementation. Often, the sample that will be examined is not based on random sampling. The first two or three subjects enrolled are usually monitored from beginning to end of each site visit. The sample to be examined and/or data to be verified varies and may include only 25% of subjects of a study that has 1,000 subjects, all subjects with a Serious Adverse Event, sites with a large number of enrollees (the FDA often reviews these data), or sites with data that are too "clean," as a certain error rate is normal and expected.
- Monitoring reports indicate significant findings from site visits, including noncompliance with protocol procedures and problems with regulatory documents (e.g., IRB reviews).
- Looking at data quality monitoring to assess GCP compliance may be challenging (e.g., CRFs are study specific).
- Monitoring plans for a given site often change.
- The monitoring of a site can either increase or decrease based on recommendations in the site monitoring reports.

Site Monitoring Templates

- Templates are available for site assessment visits, site initiation visits, interim monitoring visits, assessment visits (site operations, pharmacy, and regulatory), and closeout visits (protocol and site).
- Forms vary based on the type of trial (e.g., treatment, vaccine, and non-drug trials).
- One Division is in the process of reevaluating and re-standardizing forms so that they can be used across their programs with the understanding that on some forms certain items may not apply.
- Monitors use several checklists, depending on the protocol and Division branches.

- For the most part, site monitoring templates are accessible (they are posted on the Web site).

GCP Compliance Associated with the GCP CBT Program Is Challenging to Evaluate

A consensus across interviewees was the challenge to get reliable, high-quality measurements from site monitoring visits regarding applicability of GCP in the conduct of clinical trial research. GCP compliance assessed from site monitoring cannot be associated with GCP CBT for various reasons:

1. There are too many staff involved in the site monitoring process.
2. It is impossible to monitor everything in every visit.
3. There are too many staff involved in the conduct of clinical trials (e.g., PI, clinical physicians, study coordinator, nurses).
4. The content of the monitoring process is consistent, but variability in aspects such as (a) the number of people involved (b) different levels of monitoring because of study characteristics (e.g., design), and (c) issues when conducting research—which are often out of the researchers' control (e.g., staff changes or delays to initiate a study due to a local IRB, both of which do not imply an incorrect implementation of GCP)—are likely to influence a fair assessment of GCP compliance associated with GCP CBT.

   Assessing GCP compliance is further complicated by the varied nature and characteristics of a study; for example, when a study protocol is more complex than usual, having more study-specific procedures to follow or being conducted at an inpatient facility on more severely ill study participants. It is likely in this case that there will be more findings cited in the monitoring process than in an observation or prevention study with healthy study subjects which would likely yield fewer Adverse Events (AEs) and Serious Adverse Events (SAEs).
5. Violations can happen because of protocol or study participants. Problems with subject behavior and noncompliance issues are likely to affect and skew data.
6. Training issues will affect the validity and reliability of data collected. Often, research staff have some knowledge of GCP already. Site monitors conduct a substantial amount of GCP training during their site visits.
7. Timelines (e.g., from research award to startup, from study startup to completion) are not appropriate to assess GCP use because there are different timelines for different parts of a research study.

Monitoring Reports and Related Documentation on GCP Compliance

Interviewees provided information on existing documentation on GCP compliance:

- Site monitoring reports
- Trigger summary reports (Extranet: It is available and at-a-glance would provide an overall assessment of the site's performance based on predetermined key indicators.)
- Quarterly progress report metrics
- Information provided to various site evaluation committees (e.g., SES, NES), which would show improvement in site performance over time due to external factors such as GCP training.

However, access to existing documentation on GCP compliance for the purpose of evaluation can be difficult. According to interviewees:

- Gaining access to sites' reports and related documentation may be a challenge because of concerns about confidentiality and disclosure of information (OMB clearance). Most site monitoring-related reports are for internal use only.
- Although the Evaluation Contractor could request access to some of that documentation for review, the extent to which NIAID monitoring staff would be open or responsive to those requests is unknown. Even if identifiable information in existing records was "whited out," it would be laborious to do it.

A General Perception of GCP Compliance Over Time ("gut feeling")

Interviewees agreed that monitors would have a general perception of GCP compliance and progress over time of a given study (a "gut feeling").

- Monitors are likely to have a general perception of GCP adherence from various indicators such as looking at a site's corrective actions to issues identified during monitoring, staff turnover, staff overwork, dynamics among research staff, relationship between PI and research staff, and credentials of study coordinators (e.g., there is a difference between having an experienced nurse or a graduate student as coordinator) or other research staff. Other indicators mentioned related to the study documentation, including organization of documentation, adequate record keeping (e.g., filing of different versions of study documents, keeping track of all documentation), and the ability to keep documentation up to date.

  A case example: monitors do not typically verify licenses of health care professionals working at a specific site. However, during the course of a regulatory file review, the monitor noticed that licenses were on file for all but one research nurse. The monitor questioned it and further investigation revealed that this nurse was in fact working in that capacity without an active license.

  Another example is when the monitor notices that site staff consistently make disparaging comments about the PI "never being around" or not really having his or her "finger on the pulse" of the day-to-day activities of the research. A prudent monitor will scrutinize study documents for signs that the PI is not "personally supervising the conduct of the study" as per obligations required on the 1572.

  Over time, a monitor might notice significant staff turnover in a given study. The monitor will seek to find if there is any link between high staff turnover and the consistency, or lack thereof, in study-specific requirements and research documentation.
- A short survey with questions can query monitors about their perceptions of GCP compliance at a given site (their "gut feeling").
- Program officers and site managers will also have a general perception of GCP use by a given study because of the monitoring process (e.g., monitor reports go to the site managers, who then inform program officers about issues in the report and current status. Site managers oversee the grant and they communicate with the onsite program officer for correction or resolution).

Key GCP Compliance Elements to Assess

Given the variety of studies conducted by NIAID's Divisions and programs, and specific study characteristics, it was of interest to learn about key elements or indicators from the site monitoring process which could inform about behavioral progress or change in GCP use that could be assessed across most or all NIAID-funded clinical trials. Interviewees suggested the following key elements:

- Decrease in number of enrollment violations, consent violations, and protocol deviations.
- Informed consent documentation (e.g., if it is signed and dated).
- Time to data entry (e.g., daily as specified, instead of weekly).
- IND (Investigational New Drug) documentation (in order and properly documented)
- Proper storage of medications.
- Checklist for inclusion and exclusion criteria.
- Report of AEs and SAEs; Divisions may vary in the way it is reported, but not *whether* it is reported, as it is a standard requirement.
- Increase in compliance with timelines for reporting AEs, EAEs, and/or SAEs.
- Decrease in number of study subjects dropped from the study or lost to follow-up if screening and consent procedures are enhanced as a result of training.
- Follow-up of protocol procedures. For example, if certifications are needed for procedures, monitoring staff check that the appropriate certifications have been obtained, or if a protocol specifies how a given drug is to be administered and research staff attended a related training, monitoring staff check that research staff administer the drug exactly as taught in the training.
- Accountability of study drugs (e.g., existence of accountability log in the pharmacy).
- How well the investigators are following the protocol.
- Use of updated forms.
- Discrepancies between reports on FDA regulation audits and site monitoring reports (need to be consistent).
- Compliance with monitoring source documentation.
- Compliance with monitoring regulatory files.
- Verification of safety reporting source data.

Veterans Affairs Site Monitoring Tool

The NOVA evaluation team identified a monitoring tool from the Veterans Affairs Good Clinical Practices Site Review Report (Sather et al., 2003)—the Veterans Affairs Site Monitoring Tool—to assess GCP compliance and queried NIAID interviewees for their comments and suggestions:

- It is a good basic checklist that provides a general assessment of GCP compliance across most clinical trial study parameters (e.g., across Divisions, types of studies).
- Its use may be perfunctory and checks may be noted without thorough review.
- It must be tailored to fit the needs of the evaluation (e.g., more space is needed for comments; it needs to be shortened to minimize burden among site monitors), yet captures essential information.
- Any Site Monitoring Tool to be completed by monitors needs to be brief, be easy to use, and capture essential information valuable to the evaluation.
- As part of a pilot study, monitors could use the form for 1 month and then provide feedback.

Additional Suggestions

Interviewees provided additional suggestions to consider when assessing GCP change over time:

- It is better to follow new studies (after GCP CBT) than ongoing or past studies. Looking at retrospective data (completed or ongoing studies) is likely to be complicated (e.g., OMB clearance) and confound findings on GCP use (e.g., studies may be at different levels of completion; those closer to completion are more likely to show GCP compliance).
- It is important to follow studies longitudinally over a reasonably long period of time. Longitudinal data can take into consideration the complexity of studies which influences GCP compliance.
- Assessment of GCP compliance can expand its focus progressively. Assessment can focus on a type of study, monitoring group, or program and expand its reach accordingly.
- It is important to collaborate and coordinate the evaluation with NIAID-funded clinical trials, research Divisions, programs, and monitoring staff.