THE GALLUP ORGANIZATION

for

National Institute on Alcohol Abuse and Alcoholism

Alcohol Research Center Program Evaluation Design

Submitted to:

NATIONAL INSTITUTE ON ALCOHOL ABUSE AND ALCOHOLISM
Planning and Evaluation Branch
Willco Building, Suite 414
6000 Executive Boulevard
Bethesda, MD 20892-7003

August 15, 2002

Submitted by:

THE GALLUP ORGANIZATION
901 F Street, NW
Washington, DC 20004

1. PROGRAM TO BE EVALUATED

1.1 Program to Be Evaluated: The Alcohol Research Centers Program

Public Law 94-371 authorized the National Alcohol Research Centers (ARC) program for the purpose of interdisciplinary research related to alcoholism and other alcohol problems. The ARCs provide:

> …long-term, typically five years, support for interdisciplinary research that focuses on particular aspects of alcohol abuse, alcoholism, or other alcohol-related problems. This program encourages outstanding scientists from many disciplines to provide a full range of expertise, approaches, and advanced technologies for developing knowledge in these areas. A primary goal of the NIAAA-funded Center is to become, through excellence in scientific research, a significant regional or national research resource. In addition, each ARC affords research training opportunities for persons from various disciplines and professions.

Currently there are 15 funded ARCs (Appendix A). ARCs are funded over a five-year period and must reapply for renewals after that interval. Length of funding among the current centers ranges from one year to over 20 years for the longest funded center.

Each ARC has four components:

- Administrative core component – an organizational structure promoting synergy among program elements
- Scientific core component – shared research resources
- Research component – interrelated scientific research projects
- Pilot project component – a flexible mechanism for developing new research activities

1.2 Program Goals

The long-term program goals or intended effects of the ARC program are defined by legislation: to provide a sustained national resource and excellence in research.

2. NEED FOR AN EVALUATION

1.2 Type of Evaluation

The purpose of this evaluation is to assess the research productivity of ARCs; the quality and merit of research conducted in ARCs; the advantages and disadvantages of specific center mechanisms, such as Core-Only Centers, Developmental Centers, Specialized Centers, and Comprehensive Centers; and the value added of having a Center rather than independent research projects at a site.

A control group will be used to compare with the ARCs. This group will consist of clusters of NIAAA-funded R01 grants. Institutions where clusters occur will be approached to participate as controls for the evaluation. Data collection on cluster institutions and individual investigators from cluster institutions will be similar to data collection from the ARCs. In addition to bibliometric outcome measures and ability to secure funding, ARCs and controls will be compared on promotion of collaborative efforts between or among projects that are scientifically related, and the extent of shared resources or facilities, institutional support, and recognition.

2.2 Purpose of the Evaluation

The National Advisory Council of NIAAA requested a review of the ARC program in order to identify potential improvement opportunities for the program, strengthen the national capacity to perform alcohol related research, and assess the productivity of the ARCs.

NIAAA is interested in answering the following evaluation questions concerning the ARCs:

1. Have the ARCs been scientifically productive?
2. Are they centers of excellence for alcohol research?
3. What is the value added of having a center?

The utility of any National Institute of Health (NIH) funding program must be evaluated in terms of its contribution to biomedical science. Productivity for individual investigators has long been measured by choosing relevant and available indicators of productive activities in the cycle of research: obtaining research funding, doing the research, disseminating the results, relying on the merit of the work to secure continued funding.

NIAAA would also like to determine whether the P50 mechanism is an effective way to promote research on alcohol. First, is a specialized center grant (P50), such as the ARC program, superior to a cluster of investigator-initiated awards (R01s), that is, is a center as scientifically productive? Second, do centers exhibit levels of quality, productivity, and innovation that are nationally and internationally recognized? Third, does a p50 award place the awardee institution on a trajectory of productivity that is not observed in institutions not awarded a P50? These questions may be answered by developing an evaluation design that compares research process indicators of the four types noted above between Alcohol Research centers and 1) institutions with clusters of R01 awards and 2) institutions that applied for P50 awards, but were unsuccessful. ARC institutions and control institutions of the two types will be compared on indicators of obtaining research funding, doing the research, disseminating the results, and the merit of the work.

2.3 Use of Results

In addition to developing an approach for monitoring and assessing the progress of the ARC program, the results of the evaluation will be useful to researchers, administrators, and academic institutions interested in strengthening their capacity to do multidisciplinary research and other researchers interested in assessing the productivity of research. Both the National Institute of Dental and Craniofacial Research as well as the National Institute of Child Health and Human Development are interested in the results of this research. These institutes, as well as others who are interested, will form a working group to address common issues in the evaluation of NIH research centers.

Measuring the outcome from research is one of the most difficult tasks facing NIH, and according to the National Academy of Sciences, Committee on Science, Engineering, and Public Policy (COSEPUP), the most crucial for the long-term health of the nation. We must find ways to ensure that basic research programs funded by the nation generate the kinds of knowledge that give us great practical benefits.[1] The results of the present study will be shared with co-sponsoring institutes.

2.4 Review of the Literature

This section summarizes previous evaluations of research center grants funded by National Institutes of Health and the National Science Foundation (NSF). The purpose of this review is to identify approaches such as evaluation designs, measures, and comparison groups used in these earlier studies. The earliest of the studies were reported in 1978 and the most recent in 2000. Sponsoring organizations are distributed as follows:

- National Center for Research Resources (NCRR) – 5
- National Institute of Dental Research (NIDR) – 2
- National Science Foundation – 2
- National Heart, Lung, and Blood Institute (NHLBI) – 1
- National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMSD) – 1
- National Institutes of Health (general) – 1

In the past decade, the National Center for Research Resources has been a leader in evaluations of research centers supporting evaluations of the Regional Primate Research Centers in 1994 and 2000, the Research Centers in Minority Institutions (RCMI) in 1995 and 1999, the General Clinical Research Centers, and the Research Infrastructure in Minority Institutions program. The National Institute of Dental Research completed two

1 Committee on Science, Engineering, and Public Policy (COSEPUP). *Evaluating Federal Research Programs*. Washington, DC: National Academy Press, 1999.

of the earliest evaluations of research centers in 1982 and 1984 and produced an important summary of center evaluations to date in 1994.[2]

Table 1 shows eight center evaluations reviewed in the 1994 paper by the NIDR Oversight Committee. Four of the studies had been completed and four were then underway. These have since been completed and more complete information is available on their results than was available to the authors of the NIDR review. Table 2 shows four additional evaluations completed by NIH institutes or centers since 1994.

2.4.1 Prior Evaluations

Output Measures: The authors of the NIDR review felt there was little controversy surrounding appropriate output measures. The evidence from the four completed evaluations at the time of their review shows all four used numbers of publications, and in most cases, the journals in which they were published, as a measure of research output. Three of those also used citation impact to measure scientific worth of those publications. Two of the three NIH-funded evaluations also measured the level of funding the centers were able to secure.

Unique among these evaluations, the evaluation of the National Science Foundation's Materials Research Laboratory Program used a panel of peer reviewers to rate 700 research publications on their "orientation and degree of innovation and integration.[3]"

Comparison Groups: Evaluation rests on the judgment of whether or not a program or project has met or exceeded expected outcomes or, alternatively, failed to do so. Determining expected outcomes is seldom simple. How many publications should a research center produce? The NIDR review authors were especially concerned with identifying comparison groups used in past and contemporary evaluations. In the case of comparison groups, significant disagreement existed as to the best approach.

2 NIDR Oversight Committee, *Background Paper on Evaluating the NIDR Centers Program*. Washington, DC: National Institute on Dental Research, 1994.

3 Ling JG, DeBolt MA, Lehl MT et al. *Evaluative Study of the Materials Research Laboratory Program*. Washington, DC: National Science Foundation Division of Materials Research, 1978. (NTIS Publication No. PB-289-867)

TABLE 1. SUMMARY OF CENTER EVALUATIONS ADAPTED FROM NIDR OVERSIGHT COMMITTEE, 1994

AGENCY OR NIH INSTITUTE OR CENTER	AUTHOR	YEAR OF REPORT	TYPE OF LARGE GRANT	COMPARISON GROUP(S)	OUTPUT MEASURES
COMPLETED EVALUATIONS					
NIH	Carter et al.	1978	P50, P30, P01	R01s	• Publications and citations • Grant activity
NSF	Ling et al.	1978	Materials Research Laboratory Program (core grants)	Universities w/out MRLs but with generous support of materials research mostly by project grants	• Publications reviewed • Citation analysis • Capabilities rated by expert panel
NIDR	Brodsky et al.	1982	Dental Research Institutes and Centers (P50)		• Publications
NIDR	Reisher and Narin	1984	Dental Research Institutes and Centers (P50)	• R01s • Mixed P50s and R01s • Before and after P50	• Publications and citations • Grant activity
ONGOING EVALUATIONS (IN 1994)					
NHLBI	Roth	1994	Programs of Excellence in Molecular Biology (P01)	Unsuccessful P01 Applicants	• Grant activity • Publications • Interviews • Destination of trainees
NCRR	Kaplan and Schindler	1996	General Clinical Research Centers (M01)	• R01s • Non-GCRC clinical investigators	• Surveys • Publications • Grant activities
NCRR	Brown	1994	Primate Research Centers (P51)	No comparison group	• Undecided
NSF	Schindel (Stine)	1996	Science and Technology Centers	No comparison group	• Publications and citations • Educational impacts • Graduates • Knowledge transfer • Patents

Source: NIDR Oversight Committee, *Background Paper on Evaluating the NIDR Centers Program*. Washington, DC: National Institute on Dental Research, 1994.

For example, a comprehensive evaluation of NIH center grants compared a group of P50, P30, and P01 grants with R01 grants.[4] The intent was to address allegations that investigators in large grants lacked the same productivity and produced science of less value than investigators supported by R01s. At the same time, some critics argued that the comparison was inappropriate as the two types of support were too different to be compared. Interestingly, large grant researchers had higher citation counts but lower funding success rates and priority scores than R01 researchers. A later study of NIDR centers found similar rates of publications per year and citations. Center grant researchers were more likely to publish in basic science journals whereas R01 researchers were more likely to publish in dental journals.

Other comparison group approaches include comparison of centers with other large grant recipients, for example, center grant researchers funded by other institutes. Both P01 and P50 mechanisms are suggested. Another approach is simply to compare outcome measures among the recipients of specialized center grants.

Another common approach is to use a before-and-after design, comparing outcomes of the programs to themselves before the center funding. Historical controls have the advantage of being similar to themselves on important dimensions that might invalidate comparisons with other institutions or investigators. However, historical comparisons may confound the effects of funding with developments that might have occurred in any case.

Finally, an approach given credence by the NIDR review is to use unsuccessful applicants for the center grant as a comparison group. They cite the National Heart, Lung, and Blood Institute's evaluation of the Programs of Excellence in Molecular Biology. That study proposed to use unsuccessful applicants based on the idea that these institutions would be closest "to the study group in research interests, goals, structure and critical mass of investigators.[5]" In contrast, a recent evaluation of the National Center for Research Resources' Research Infrastructure in Minority Institutions Program rejected using unsuccessful applicants because a group of peer reviewers had determined that they had less potential for being productive based on their track record.[6] It seems that this approach would be valid only if control institutions were limited to those considered technically acceptable with a priority score near the fundable line. The presumption is that only the limitation of funds, and not lack of potential for success as a center, kept these institutions from being funded.

---

4 Carter GM, Lai CS, Lee CL. *A Comparison of Large Grants and Research Project Grants Awarded by the National Institutes of Health*. Washington, DC: National Institutes of Health, Contract No. N01-RR-02132, 1978.

5 Roth, CA. *Evaluation of the National Heart, Lung, and Blood Institute Programs of Excellence in Molecular Biology*. Presented in "Request for Delivery Order" to Deputy Assistant Secretary for Health (Planning and Evaluation), Washington, DC: Department of Health and Human Services, 1994.

6 Wells JA, Karr D, Zimmerly M et al. (Center for Health Policy Studies). *Midcourse Assessment of the Research Infrastructure in Minority Institutions Program*. Washington, DC: National Center for Research Resources, Contract N01-OD-7-2117, Task Order 9, January 2000.

2.4.2 Recent Evaluations

Table 2 summarizes four recent center evaluations. Three of these were supported by the National Center for Research Resources and one by the National Institute of Arthritis and Musculoskeletal and Skin Diseases.

Output Measures: As in the earlier evaluations, measuring publications and citation impact are the most commonly used indicators of productivity. Similarly, levels of NIH funding and collaborative efforts among investigators are measured.

More emphasis is found in these evaluations for measurement of the number and disposition of trainees, and a new measure appears—utilization of core resources. Program trainees and graduates are seen as an important output of these programs. This measure appeared in none of the completed evaluations reviewed by NIDR, but was a planned measure in three of the ongoing evaluations. Also, utilization of core resources and facilities is measured in three of the recent evaluations. Not surprisingly, these are centers supported by NCRR, but these measures of core resources are applicable to most large grant mechanisms.

In addition, one evaluation each measured knowledge transfer and cost effectiveness. The Research Centers in Minority Institutions program has a component requiring knowledge transfer in the form of educational outreach to professionals and the community.[7] The regional Primate Research Centers program is assessed in terms of outputs per dollar spent.[8]

Comparison Groups: In general, there are no new innovations in the approach that more recent evaluations have taken toward comparison groups. However, there is more emphasis on matching case and comparison programs, researchers, or institutions on variables that make the contrast more valid. For example, both the Research Centers in Minority Institutions (RCMI) and the Research Infrastructure in Minority Institutions (RIMI) evaluations matched comparison institutions on level of PHS funding, institutional characteristics (either public private status or Carnegie classification), and location. The RCMI evaluation also matched on institutional assets.

---

7 QRC Division of Macro International. *Evaluation of the Research Centers in Minority Institutions, Final Report*. Washington, DC: National Center for Research Resources, Contract N01-OD-7-2114, Task Order 2, no date.

8 James Bell Associates. *Full Scale Evaluation of the Regional Primate Research Programs, Final Report*. Washington, DC: National Center for Research Resources, Contract N01-OD-7-2115, Task Order 3, July 2000

TABLE 2. SUMMARY OF CENTER EVALUATIONS COMPLETED SINCE NIDR OVERSIGHT COMMITTEE, 1994

AGENCY OR NIH INSTITUTE OR CENTER	AUTHOR	YEAR OF REPORT	TYPE OF LARGE GRANT	COMPARISON GROUP(S)	OUTPUT MEASURES
NIAMS	Rich	1996	NIAMS Centers (P60)	Non center-affiliated investigators with NIAMS funding > $500K/yr	• Qualitative survey of center advantages/disadvantages • Grant activity
NCRR	QRC	1999	Research Centers in Minority Institutions (G12)	Matched institutions with similar PHS research funding, public/private, assets, location	• Publications and citations • Grant activity • Graduates and fellows • Knowledge transfer • Resources and facilities
NCRR	Wells et al.	2000	Research Infrastructure in Minority Institutions (P20)	Matched institutions with similar PHS research funding, Carnegie classification, Census region	• Publications • Grant activity • Majors and graduates • Collaborative activities
NCRR	James Bell Associates	2000	Regional Primate Research Centers (P51)	Center-center comparisons Other researchers publishing on non-human primates	• Publications and citations • Trainees • Animal production and income generated • Cost-effectiveness

2.5 Timeliness of the Evaluation

The National Advisory Council of NIAAA requested a review of the ARC program in order to identify potential improvement opportunities for the program, strengthen the national capacity to perform alcohol related research, and assess the productivity of the ARCs. Some ARCs have been funded for 20 years or more without an evaluation of the program overall.

3. EVALUATION DESIGN

3.1 Study Questions

The evaluation questions are as follows:

1. What is the productivity of the ARC program as measured by bibliometric techniques, other new methodologies beyond bibliometrics, and expert reviews?

2. Have ARCs stimulated innovative new research?

3. How effective is the research infrastructure of the ARCs as measured by proficiency in obtaining grants, internal sharing or resources and knowledge gained, collaborations with other academic, State, and community institutions, and commitment form the parent institution in the form of increased resources?

4. What is the value added of having an ARC versus a number of R01s in a site?

5. What are the specific goals of NIAAA that each ARC is addressing? Are they accomplishing what they were established and mandated to do?

6. Are ARCs able to attract high quality new (both young and/or established investigators) compared with clusters of R01s?

3.2 Target Population

3.2.1 ARC Selection Criteria

There are currently 15 funded ARCs (Appendix A). However, some have been only recently funded. An interval of less than five years scarcely includes enough time for research to move from start-up to results to publication to impact of publication. Therefore, the ARC group will be limited to those that have been in existence at least five years.

Members of the treatment group (the ARCs) could serve as historical controls for themselves. The idea is that the treatment group, before the treatment is applied, should reflect what this group would be like in the absence of treatment. The assumption is that even though the control group is measured earlier in historical time, it reflects what the treatment group would be like without the treatment, even at a later date.

3.2.2 ARC Recruitment

Principal investigators of the ARCs will be contacted and asked to participate in the evaluation. Part of the evaluation design concerns data collection from individual investigators in the ARCs (and control groups). Participation for individual investigators will be solicited at the time of the data collection and will be confidential.

3.2.3 Overview of Control Institutions

3.2.3.1 R01 Clusters

This control group consists of institutions in which there is a cluster of NIAAA-funded R01 awards. These are individual investigators who are supported by NIAAA, but without the benefit of an ARC. Whether they are the beneficiaries of some other supportive arrangement is an empirical question that will be investigated in the course of the evaluation. This type of control group is normative, meaning they represent a norm or ideal of the individual investigator pursuing investigator-initiated research. There is a belief held by some that investigators in funded research centers are less productive than R01-funded investigators in the same fields. This control group provides a direct test of that notion. Moreover, there is interest in understanding the value added by the organizational arrangements implied by a research center. This comparison will address those issues as well.

3.2.3.2 ARC Applicants

This control group consists of unsuccessful applicants for an ARC award. When contrasting this group with the successful ARC applicants, one is able to address questions of productivity and quality/impact. The applicants get selected not solely as a function of the quality of their application. There are so many factors that contribute to "error variance", the non-ARC applicant group comes the closest of the three comparison groups to holding all "else equal." Baseline measures of productivity and impact will be obtained and compared to successful applicants.

Using unsuccessful applicants as controls provides for a treatment-control design that has outward similarity to a true experimental design in which the control group entities were eligible to receive the treatment, but did not only due to chance (i.e., the process of randomization). In the absence or randomization, it is important to admit to the control group only those entities that would have been eligible to receive the treatment. Thus, in the case of Alcohol Treatment Centers, the controls should be institutions that have the prerequisites for being an ARC. This is only true, however, if they meet the eligibility

criteria for funding. One would not want to admit to the controls an institution that reviewers found unqualified in a major way. The presumption must be that the controls differ from the ARCs in no significant way that is related to scientific productivity, but only in incidental ways related to accidents of the application process, reviewer proclivities, and "luck". It is easy to argue for the comparability of the institution immediately above the funding line and that immediately below, but it is an empirical questions whether there are sufficient comparable institutions below the funding line.

Similarly, whether the eligible controls are comparable on factors known to be related to the outcomes is open to empirical test, where the data exist for doing so. After the control group is assembled, statistical comparisons can be made on factors potentially related to the outcomes under study. A more controlled approach would consist in matching important traits between experimental and control entities, but the availability of data for doing so is questionable.

3.2.3.3 Historical Controls

It is recommended that historical control observations be collected for both types of control institutions. Figure 1 (next page) shows a representation of the type of comparable control group proposed for the evaluation. The T is treatment and the Os represent measurements on the outcomes (performance measures) at times before and after the treatment intervention. The feasibility of the design may vary for different outcome measures. For example, citation indexes exist for all periods, but trainee information, which relies on reports of the institutions, may be difficult to find on a consistent basis for years past. Measurement may have to be limited to one or two five-year periods.

Figure 1. Comparable Control Design

Treatment (Alcohol Research Centers)	O_{-1}	T	O_1	O_2	O_3
Control (R01 Clusters)	O_{-1}		O_1	O_2	O_3
Control (ARC Applicants)	O_{-1}		O_1	O_2	O_3

3.2.4 Eligibility and Recruitment of Control Institutions

3.2.4.1 R01 Clusters.

Eligibility criteria for a cluster of R01s are the existence of at least five R01 projects funded by NIAAA with at least four separate principal investigators in the most recent five-year period. Furthermore, at least two principal investigators must have their primary appointment in the same department. The rationale for these criteria is that a typical Alcohol Research Center award comprises four or five research components supporting a similar number of investigators. These are roughly comparable to an

independent, investigator-initiated research project. It is also typical for some of these investigators to have appointments in the same department.

To assess the feasibility of these criteria, the feasibility study contractor selected ten medical schools at random from the American Association of Medical Colleges' (AAMC) list of 122 schools in the United States (excluding Puerto Rico)[9]. Of these, five met the criteria based on NIAAA-funded research awards reported in CRISP for these institutions between 1998 and 2002. The number of awards averaged about 6, ranging from zero (two institutions) to fourteen (also two institutions). The number of investigators averaged about five and ranged from zero to thirteen. This implies that in the population of medical schools, about 61 institutions would meet the criteria. Moreover, there is a 95 percent probability, based on this sample, that the true proportion of eligible schools is greater than 35 percent, minimally 42 schools. Additionally, there may be clusters identified in institutions without medical schools. This should provide ample opportunity to recruit 15 institutions for the evaluation. Statistically, two matched controls provide significantly greater precision than one control per ARC, but the number of available clusters meeting a reasonable size criterion will be a limiting factor. This type of control group is normative, meaning they represent a norm or ideal of the individual investigator pursuing investigator-initiated research. There is a belief held by some that investigators in funded research centers are less productive than R01-funded investigators in the same fields. This control group provides a direct test of that notion. Moreover, there is interest in understanding the value added by the organizational arrangements implied by a research center. This comparison will address those issues as well.

The R01 recipients should be matched by zero time, that is, they should be selected from among awardees in the same periods in which ARCs were initiated. Since an individual may have multiple awards, he or she should be eligible for selection in each period in which they received an award. Clusters will be evaluated for a five-year period comparable to the five-year periods of ARC funding.

Institutions with identified clusters of NIAAA-funded R01s will be contacted and asked to participate in the evaluation. They will be asked to identify someone who would serve as the primary contact for purposes of the evaluation. This might be one of the R01 principal investigators or the chair of the department in which most of the R01s are funded.

3.2.4.2 Unsuccessful ARC Applicants

Unsuccessful applicants will be identified from NIH records. Priority scores will be ascertained so that only institutions near the funding line may be selected. Institutions that were applicants when the ARCs applied will be pooled and a group of controls will

[9] The ten medical schools were Boston University, George Washington University, Loyola University-Chicago, Northeastern Ohio Universities, Temple University, University of Arkansas, University of Kansas, University of Nevada, University of Tennessee Health Science Center, and Virginia Commonwealth University.

be randomly selected, one per ARC. Institutions will be contacted and asked to participate in the evaluation. They will be asked to identify someone who would serve as the primary contact for purposes of the evaluation. This might be the principal investigator named in the unsuccessful application or the chair of the department in which the application was developed.

Eligibility criteria and factors differing between treatment and control groups that will affect the evaluation outcomes will be measured as close to zero time as possible. That is the time at which randomization would have occurred had the treatment and control group entities been assigned experimentally. Thus, the institutions chosen for the control group will be selected from applicants who were in the pool when the member of the ARC treatment group was funded. This does not ensure comparability, but it ensures that the institutions were at the same relative stage of development, on average.

3.3 Key Variables

The utility of any NIH funding program must be evaluated in terms of its contribution to biomedical science. Productivity for individual investigators has long been measured by choosing relevant and available indicators of productive activities in the cycle of research: obtaining research funding, doing the research, disseminating the results, relying on the merit of the work to secure continued funding, and repeating the cycle.

The principal outcome measure reflecting dissemination of research results is the number of publications. Scientific worth or quality is most commonly measured by bibliometric analysis, usually citation impact. When evaluations have tracked the ability of researchers to use their results to secure additional research funding, the measure has been the number and dollar amount of awards, especially investigator-initiated research (R01) awards. Research performance, and how the research is done, is seldom measured although a few evaluations have measured collaborative publication. The present evaluation hopes to go beyond the usual measures such as bibliometrics wherever this is feasible.

Table 3 summarizes the concepts of the research cycle, indicators of each concept, and derived variables used in the evaluation analysis. Operationalization of the variables is described the sections that follow.

4. DATA COLLECTION AND ANALYSIS

4.1 Data Sources

It is anticipated that measures will be collected from the following data sources as identified in Appendix B.

Table 3. Scientific Research Cycle Concepts, Indicators, and Variables

Concept	Indicator	Variables
Research Support	Awards (Grants and Cooperative Agreements)	Total Award Amount Award Amount per Year Award Amount per Investigator Awards from NIAAA versus Other NIH versus Other Sources R01 Awards Research Project Awards Other Awards (by Activity Code)
Research Performance	Core Resources/ Technology Transfer	Count of Uses of Core Resources Average Use of Core Resources per Year Number of Technology Transfer Requests
	Collaboration	Count of Collaborative Projects Proportion of Investigators Involved in Collaborative Projects Count of Collaborative Papers Average Number of Collaborative Papers per Year Average Number of Collaborative Papers per Investigator
Dissemination of Results	Publications	Number of Publications Average Publications per Year Average Publications per Investigator Number of Publications per Journal
	Training	Number of Trainees Average Trainees per Year Proportion of Trainees Going to Institution with an ARC versus Control Institutions versus Other
Scientific Worth	Bibliometrics	Total Citations Average Citations per Year Average Citations per Investigator Citation Impact Average Citation Impact per Year Average Citation Impact per Investigator Average Citation Impact of Journals
	Honors/Awards	Total Honors/Awards Honors/Awards per Year
	Technology Transfer	Diagnostic Tests, Questionnaires, Animal Models, Patents, etc.

- NIAAA Administrative records — By ARC and control institution, principal investigator; award or requested amount; and key personnel.

- IMPAC II; NSF Awards Database — By principal investigator, the dates; duration; amount; level of support; and institutional sponsor of all awards.

- Citation Database — By key personnel, the number; dates; journals; and citation impacts of all publications in refereed journals. Source may be ISI or the SPIRES system if it is ready by implementation of the evaluation.

- Investigator Interview — From ARC and control group key personnel, utilization of core scientific resources in research; participation in technology transfer and knowledge transfer activities; honors and awards; and invited presentations.

- Administrative Questionnaire — From principal investigators/administrators of ARCs, technology transfer activities; number of trainees; destination of trainees.

4.2 Data Collection Strategies

Three principal data collections strategies will be employed in this evaluation. First, NIH documents will be reviewed. These include initial applications for the Alcohol Research Center (P50) grants, progress reports, and competing continuation applications. In addition, training grant (T32) applications and progress reports, and R01 applications and progress reports will be reviewed. The documents will be mined as sources of counts of activities such as number of trainees and utilization of core resources.

Second, data will be abstracted from existing databases. These include the NIH IMPAC database to acquire information on research and training grant awards and citation indexes such as the ISI *Web of Science* database or SPIRES. NIH is currently developing the SPIRES system that links publications in Medline and awards in IMPAC. The feasibility of using SPIRES is being explored by NIAAA.

Third, data will be acquired from questionnaires administered to alcohol researchers in ARCs and control institutions.

4.3 Data Collection Instruments

There are two data collection instruments. Draft documents are shown in Appendix C.

- Investigator Interview — From ARC and control group key personnel, utilization of core scientific resources in research; participation in technology transfer and knowledge transfer activities; honors and awards; and invited presentations.

- Administrative Questionnaire — From principal investigators/administrators of ARCs, technology transfer activities; number of trainees; destination of trainees.

4.4 Clearance Requirements

The data collection instruments will be completed by from 45 to 90 individuals. The interview is semi-structured and may not require OMB clearance. The administrative survey is designed to collect statistical data and will require OMB clearance. When the contractor has developed final versions of the instruments, NIAAA will submit an OMB clearance package.

The instruments will be developed to minimize burden and will not contain any questions of a sensitive nature.

4.5 Data Integrity

Applicants for the evaluation contract will be required to explain how standards for data integrity will be developed and maintained. At the very least, contract proposals will be required to include a detailed plan that explains the following:

- Protocols for selecting individuals to complete the data collection instruments
- A coding scheme for the resulting data
- Reliability checks of the accuracy of data coding and entry

Data integrity concerns are also pertinent to archival data. Reliability checks for the various data sources will depend on the type of data secured. For instance, reliability checks of the effectiveness of an abstract of grant awards from IMPAC may be compared with reports in center applications and individual investigators' curricula vitae. Again, proposals will be required to include a detailed plan for ensuring the reliability of all archival data.

4.6 Ethical Considerations

The nature of the initiative imposes demands on the evaluator to ensure that the anonymity of institutions with ARCs and those selected as controls be protected in all reports and other deliverables that may be seen by the public. It will be essential not only to protect the confidentiality of institutions and informants, but also to ensure that no information be included in any reports of individual site visits that might assist an informed observer in identifying any institution or individual. The contractor will be required to submit a plan that specifies the various steps that will be taken to ensure the confidentiality of all data. As an added protection, the evaluator's Institutional Review Board must approve the evaluation plan for the protection of human subjects.

4.7 Data Preparation

Data will be combined into two databases—the Center-Level Database and the Investigator-Level Database. The Center-Level Database will include data aggregated to the level of the institution, including those with ARCs as well as control institutions. Thus it will include both data that can only be evaluated at the institutional level, such as the number of trainees per year, and data that is aggregated from investigator level outcomes, such as the average number of publications per investigator per year.

The evaluator will be required to submit a detailed data preparation plan. The plan will include the following:

- Detailed plans for coding variables collected from questionnaires. This will include production of a codebook for each data source indicating variable names, variable labels, and value labels.

- Detailed plans for processing data from archival sources. This will include production of a codebook for each data source indicating variable names, variable labels, and value labels. In addition, the codebook will discuss decision rules for inclusion of data where disagreements may be observed across sites—for example, the categorization of trainees may differ among the sites—and within sites—for example, different sources may not agree as to utilization of core resources.

- Detailed plans for combining data from diverse sources such as questionnaires, bibliometric analysis and NIH grant application archives.

- Detailed plan and file structure for the Center-Level Database and Investigator-Level Database.

In addition, the evaluator will lay out a plan for storing, classifying, and retrieving all data and ensuring the confidentiality of all data or study findings that might identify specific institutions or investigators.

4.8 Data Analysis

The analysis will include assessment of differences among evaluation groups on both quantitative and qualitative variables. For quantitative variables, the data from various sources—questionnaires, CRISP/IMPAC, citation analysis, coded NIH documents—will be merged into a working file. Initial analysis will focus on the validity and integrity of the data. These procedures will include frequency distributions, means of quantitative variables, proportions of categorical variables and correlations or cross tabulations among variables to ensure that numerical codes are within the appropriate ranges and internally consistent.

The analysis used to assess evaluation questions will use ANOVA and related statistical

techniques. There are three evaluation groups—the Alcohol Research Centers and two groups of controls shown in the table below.

Evaluation Group	Group Designation	Mean of Performance Variable	Dummy Variables	
			X_1	X_2
Alcohol Research Centers (ARCs)	A	$\overline{Y}_A$	0	0
Unsuccessful ARC Applicants	B	$\overline{Y}_B$	1	0
Clusters of R01s	C	$\overline{Y}_C$	0	1

For each group, the evaluation will have measured performance indicators designated Y. ANOVA assesses the magnitude and statistical significance of differences among the group means of the performance measure. One way to compute the needed ANOVA values is through linear regression as in the following equation.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e$$

Where Y is defined as above, X_1, and X_2 denote group membership, and e is random error.

A solution to this model will result in the following equalities:

$$\beta_0 = \overline{Y}_A,\ \beta_1 = \overline{Y}_B - \overline{Y}_A,\ \beta_2 = \overline{Y}_C - \overline{Y}_A$$

In other words, the value of β_1 is the difference in the mean of the performance measure between the Alcohol Research Centers and the control group of unsuccessful ARC applicants. Likewise, the value of β_2 is the difference in the mean of the performance measure between the Alcohol Research Centers and the control group of R01 clusters. The t-tests applied to β_1 and β_2 are the tests of the statistical significance of the differences noted above. Performance measures that are dichotomous or consist of counts may be similarly treated using logistic regression or log-linear weighted least squares.

An advantage of the regression approach is the ease with which covariates may be included in the analysis. In a non-experimental design, there may be factors that differ among the evaluation groups that could account for the differences observed in β_1 and β_2. For example, two groups could differ in the number of publications produced, but be observed to also differ in the average level of grant support. When grant support is included in the equation, the differences in publications among evaluation groups may be reduced. This part of the raw observed difference in groups will have been shown to be the results of funding differences, not of the funding vehicle per se. This approach would also accommodate time parameters as well providing the ability to model changes over time.

The analysis will assess differences among the groups on potentially confounding variables and other covariates. Those that are found to significantly differ among the groups will be tested to determine the degree to which they affect the estimates of differences among groups. Some variables will not lend themselves to quantitative analysis, such as reports of the advantages of a center, or reports of how institutions provide support for research under the varying conditions defined by the evaluation groups. These data will be gathered in semi-structured interviews. The interviews will be recorded and transcribed and at least two coders will listen to and code each interview. The two coders and the project director, acting as referee, will develop a coding scheme for the qualitative interviews. The initial coding scheme will be based on three interviewees from each evaluation group. The coders will apply the coding scheme to the test interviews and the results will be compared. Where there are systematic disagreements, the referee will discuss the results and agreement will be reached on an appropriate coding standard. This process will continue until substantial agreement in reached in coding the same interview. Data from both the quantitative and qualitative analyses will be summarized and presented in a systematic manner that addresses each of the specific evaluation questions. The evaluator will be required to submit a detailed data analysis plan. The plan will include specifications of the types of univariate and multivariate analysis and statistical tests to be used. The analysis will center on difference between ARCs and control institutions on both center level and investigator level data. The evaluator will be required to provide table shells for the analyses that will be undertaken.

The evaluator will be required to provide table shells for the analyses that will be undertaken.

Table 4 shows a sample table shell for variables related to grant awards to ARC and control institutions.

Table 4. Table Shell of ARC versus Control Site Differences on Grant Awards

Variable	ARC	Control	Significance Test
Total Award Amount			
Average Award Amount per Year			
Average Award Amount per Investigator			
Awards from NIAAA as proportion of all Awards			
Total R01 Awards Amount			
Average R01 Award Amount per Year			
Average R01 Award Amount per Investigator			
Awards from NIAAA as proportion of all R01 Awards			
Total Research Awards Amount			
Average Research Award Amount per Year			
Average Research Award Amount per Investigator			
Awards from NIAAA as proportion of all Research Awards			

5. EVALUATION RESULTS

5.1 Products of the Evaluation

- An executive summary of the purpose, methodology, key findings, and recommendations of the evaluation.

- A final report discussing specific findings.

- A report on the Delphi technique prioritizing recommended actions based on evaluation findings.

5.2 Dissemination of Results

The final report will be provided to principal investigators of ARCs and contacts at participating control institutions. The executive summary and prioritized recommendations will be disseminated to administrative staff at NIH and other pertinent

Federal agencies, members of Congress, members of the scientific community, the press, and the public.

6. PROJECT MANAGEMENT

6.1 Project Implementation

The evaluator will be required to lay out clear plans for project management, implementation, quality control, and protection of confidentiality. The credentials and experience of staff that manage the evaluation should include professionals with specific experience in the evaluation of NIH center programs. Staff who design data collection material and procedures should exhibit a proven track record in questionnaire design for research professionals and the credentials and experience of those who analyze the data should demonstrate experience on the manipulation and reporting of data. The evaluation contractor will be required to identify specific evaluation professionals, questionnaire design experts, and data analysts as key staff and to make binding commitments to their availability for the study.

There are 15 evaluation tasks comprising the evaluation. The figure in section 6.3 presents these tasks along with associated deliverables and anticipated due dates.

Task 1 – Initial Project Meeting

The contractor should immediately schedule a meeting with the Project Officer and other NIAAA staff to discuss the scope, objectives, and schedule for the project. NIAAA will prepare the agenda for the meeting and will provide materials the contractor will need for completion of the project. The contractor will provide a written summary of action items discussed at the meeting.

Task 2 – Develop Evaluation Workplan

The contractor will prepare a detailed workplan for the project discussing procedures for the completion of each task and the production of the required deliverables. The contractor will prepare a detailed timeline for deliverables and will identify and discuss any critical paths or anticipated problems.

The contractor will submit a draft workplan. NIAAA will review and comment on the draft within five working days. The contractor will then make any necessary revisions and submit the final workplan.

Task 3 – Convene First Meeting of Advisory Panel

The contractor will work with NIAAA staff to identify potential members of the advisory panel. NIAAA will determine a prioritized final list of potential advisory panel members. The contractor will contact members in order of priority and invite them to participate on the evaluation advisory panel.

Once the panel has been completely identified, the contractor will determine a date for the first meeting of the advisory panel and plan for the meeting. The contractor will provide the workplan and any other pertinent materials to the advisory panel with sufficient time to review materials before the meeting. The contractor will convene the advisory panel to discuss the role of the advisory panel and the workplan.

The contractor will provide a written summary of action items discussed at the meeting and will make any necessary changes to the evaluation workplan.

Task 4 – Identify Control Groups

The contractor will work with NIAAA to obtain access to the NIH IMPAC II database and will use this database to identify clusters of NIAAA-funded R01 grants. The contractor will submit a comprehensive list of clusters to NIAAA and will work with them to determine a minimum cluster size for eligibility in the control group. Institutions may be matched to ARCs based on the number of alcohol investigators in a given year, although matching is not required. Among possible control institutions, a control group will be randomly selected such that there are at least two control institutions per ARC. The contractor will submit the selected control institutions, lists of principal investigators, their pertinent awards, and their departmental affiliations.

Task 5 – Prepare OMB Submission

The contractor will prepare an OMB Clearance Package for the ARC evaluation. The OMB clearance package will include information about the evaluation design, selection of institutions and of individuals within institutions, data collection instruments, and estimated burden.

The contractor will submit a draft OMB Clearance Package. NIAAA will review and comment on the draft within five working days. The contractor will then make any necessary revisions and submit the final OMB Clearance Package.

Task 6 – Finalize Data Collection Instruments and Procedures

Appendix B identifies the required evaluation measures and provides protocols for their collection. Measures fall into the following four categories:

- Research Support
- Research Performance
- Dissemination of Results
- Scientific Worth

It is anticipated that measures will be collected from the following data sources as identified in Appendix B:

- NIAAA Administrative records — By ARC and ARC applicant, principal investigator; award or requested amount; and key personnel.

- IMPAC II; NSF Awards Database — By principal investigator, the dates; duration; amount; level of support; and institutional sponsor of all awards.

- Citation Database — By key personnel, the number; dates; journals; and citation impacts of all publications in refereed journals.

- Investigator Interview— From ARC and control group key personnel, utilization of core scientific resources in research; participation in technology transfer and knowledge transfer activities; honors and awards; and invited presentations.

- Administrative Questionnaire — From principal investigators/administrators of ARCs, technology transfer activities; number of trainees; destination of trainees.

The contractor will review the measurement protocols and prepare the necessary data collection instruments and procedures required to collect the evaluation measures for ARC and control institutions for the relevant years.

The contractor will submit draft data collection instruments and procedures. NIAAA will review and comment on the draft within five working days. The contractor will then make any necessary revisions and submit the final data collection instruments and procedures.

Task 7 – Review ARC Documentation

The contractor will review the documentation concerning the ARCs provided by NIAAA. At minimum, this will include the initial applications and annual renewals for each ARC. The contractor will record the required information from this data source including the names of investigators participating in each ARC over time and other pertinent information such as core services, collaborations, and products of the ARC.

Task 8 – Conduct Surveys

Surveys will be conducted of ARC principal investigators and control institution principal representatives designated for control institutions (Administrative Questionnaire) as well as investigators in ARCs and control institutions (Investigator Questionnaire, Peer Questionnaire). The contractor will disseminate the instruments using paper or Web mode with appropriate reminders and follow-up.

During the data collection, the contractor will provide weekly reports on questionnaire returns and response rates.

Once data collection is completed, the contractor will submit a summary of the survey data consisting of frequencies and summary statistics.

Task 9 – Collect Citation Data

The contractor will prepare a list of investigators from the ARCs and control institutions to submit for publication and citation analysis. The contractor will identify all peer reviewed scientific publications of these investigators for the pertinent years of the evaluation. Additionally, the contractor will determine the number of citations for each publication, citation impact per item per year, and the citation impact per year of all journals in which the investigators have published.

The contractor will submit a summary of publication and citation data to NIAAA.

Task 10 – Analyze Data

The contractor will develop an analysis plan and submit table shells to NIAAA for review. NIAAA will review and comment on the draft within five working days. The contractor will then make any necessary revisions and submit the final analysis table shells.

At minimum, the analysis will compare each of the evaluation measures between the ARC and control institutions. The contractor will provide a narrative describing any statistical or methodological limitations pertaining to the comparisons.

Task 11 – Submit Draft Final Report

The contractor will prepare and submit a draft report. The final report will include an overview of the evaluation, evaluation methodology, results of the evaluation, and conclusions and recommendations. NIAAA will review and comment on the draft within five working days. The contractor will then make any necessary revisions and submit a revised version of the draft report. This report will be disseminated to the advisory panel.

Task 12 – Convene Second Meeting of Advisory Panel

The contractor will determine a date for the second meeting of the advisory panel and plan for the meeting. The contractor will provide the draft report and any other pertinent materials to the advisory panel with sufficient time to review materials before the meeting. The contractor will convene the advisory panel to review each of the specific evaluation findings as they pertain to the evaluation questions and provide consensus conclusions and recommendations to NIAAA concerning the ARC program.

The contractor will provide a written summary of action items discussed at the meeting and will make any necessary changes to the evaluation workplan.

Task 13 – Submit Final Report

The contractor will prepare and submit a draft final report. The final report will be a revised version of the draft report incorporating any changes suggested by the advisory panel and their conclusions and recommendations. NIAAA will review and comment on the draft within five working days. The contractor will then make any necessary revisions and submit the final report.

Task 14 – Debrief NIAAA Staff

The contractor will work with NIAAA to schedule a meeting with senior executives of NIAAA and NIH as designated by NIAAA. The contractor will present the findings of the evaluation and recommendations of the advisory panel.

Task 15 – Project Management and Reporting

The contractor will conduct a regularly scheduled weekly telephone call with the Project Officer. The content of the call will include accomplishments of the past week, anticipated activity in the coming week, problems encountered or anticipated, and planned solutions to those problems.

The contractor will prepare a monthly report with similar content and submit it by the tenth day of each month during the project except the first and last months.

6.2 Advisory Committee

It is anticipated that the evaluation will employ an advisory committee. The committee will consist of four to six senior scientists from institutions other than those included in the evaluation. NIAAA staff will nominate these scientists and their participation will be solicited and confirmed by the contractor. The scientists may be in the alcohol research field or in related disciplines.

The Scientific Advisory Committee (SAC) will meet on two occasions. First, they will meet after the contractor has submitted a draft workplan for the evaluation. This meeting will be convened by week 14 of the evaluation. The SAC will review the evaluation workplan and provide feedback to responsible NIAAA staff and the contractor's project leadership regarding the evaluation design and implementation plan. Second, the SAC will meet after the contractor has submitted the draft final report. This meeting will be convened by week 66 of the evaluation. The contractor will conduct a modified Delphi technique, a method for group decision-making, to assist the SAC members and NIAAA staff to achieve consensus on evaluation conclusions and priorities for recommended actions.

6.3 Estimated Timeline for the Evaluation

It is anticipated that the evaluation will require approximately 2 years, as specified in the table on the following page.

Task Plan and Timeline: Alcohol Research Center Program Evaluation

Task	Deliverable	Date Due*
Task 1 – Initial Project Meeting	Submit Summary of Initial Project Meeting	2 weeks
Task 2 – Develop Evaluation Workplan	Submit Draft Workplan	6 weeks
	Submit Final Workplan	8 weeks
Task 3 – Convene Advisory Panel	Invite Experts to Sit on Advisory Panel	12 weeks
	Convene First Advisory Panel Meeting	20weeks
	Submit Summary of First Advisory Panel Meeting	22 weeks
Task 4 – Finalize Data Collection Instruments and Procedures	Submit Draft Data Collection Instruments	14 weeks
	Submit Final Data Collection Instruments	18 weeks
Task 5 – Prepare OMB Submission	OMB Clearance Document	20 Weeks
Task 6 – Identify Control Groups	Submit Initial List of Controls	24 weeks
	Submit Final List of Controls	28 weeks
	Submit Summary of Control Data	36 weeks
Task 7 – Review ARC Documentation	Submit Summary of ARC Data	44 weeks
Task 8 – Conduct Surveys	Submit Summary of Survey Data	64 weeks
Task 9 – Collect Citation Data	Submit Summary of Citation Data	64 weeks
Task 10 – Analyze Data	Submit Analysis Tables Shells	68 weeks
	Submit Analysis Tables	72 weeks
Task 11 – Submit Draft Report	Submit Draft Report	84 weeks
	Submit Revised Draft Report	92 weeks
Task 12 – Convene Advisory Panel	Convene Second Advisory Panel Meeting	92 weeks
	Submit Summary of Second Advisory Panel Meeting	94 weeks
Task 13 – Submit Final Report	Submit Draft Final Report	96 weeks
	Submit Final Report	102 weeks
Task 14 – Debrief NIAAA Staff	Conduct Debriefing	104 weeks
Task 15 – Project Management and Reporting	Conduct Weekly Phone Call with NIAAA	Weekly
	Submit Monthly Report	10th day of month

- Following effective date of contract.

7. BUDGET ESTIMATE

7.1 Estimated Cost

Estimated costs are as follows:

ESTIMATED COSTS

Personnel	Quantity	Rate	Cost	Total
Project Director				
Senior Project Advisor				
Survey Administrator				
Statistical Analyst				
Questionnaire Specialist				
Clerical/Transcriptionist/Coder				
On-Site Data Collectors				
Subtotal	5636			$546,730
Other Direct Costs				
Postage/Courier	78	$ 15	$ 1,170	
Computer	5,636	$ 0.79	$ 4,452	
Copying	300,000	$ 0.25	$ 75,000	
Travel*			$ 18,590	
Commercial Data Sources**	395	$ 200	$ 81,000	
Honoraria for Advisory Panel	20	$ 250	$ 5,000	
Scientific Consultants	10	$ 800	$ 8,000	
Subtotal				$193,212
G&A on Other Direct Costs				$ 42,816
Total				$782,758

* Travel detail is as follows: Airfare $500/trip, Ground $40/day, Per Diem $60/day, Lodging $145/night. Each trip is estimated for two days and one night at $845. Proposed trips include: 6 trips for two Gallup personnel to sites and 2 trips for 5 advisory committee members.

** Bibliometric analysis is $2,000 for the first five investigators and $200 per investigator thereafter. We estimate 400 investigators.

7.2 Anticipated Funding Sources

The anticipated source of funding for this evaluation is NIH's 1% evaluation set-aside funds

Appendix A

NIAAA Alcohol Research Centers

Indiana University School of Medicine
Indianapolis, Indiana

Louisiana State University Health Sciences Center
New Orleans, Louisiana

Medical University of South Carolina
Charleston, South Carolina

Oregon Health Sciences University & Department of Veterans Affairs Medical Center
Portland, Oregon

Pacific Institute for Research and Evaluation
Berkeley, California

Public Health Institute
Berkeley, California

The Scripps Research Institute
La Jolla, California

Thomas Jefferson University
Philadelphia, Pennsylvania

University of Colorado Health Sciences Center
Denver and Boulder, Colorado

University of Connecticut Health Center
Farmington, Connecticut

University of North Carolina
Chapel Hill, North Carolina

University of Pittsburgh School of Medicine
Pittsburgh, Pennsylvania

University of Southern California
Los Angeles, California

Wake Forest University
Winston-Salem, North Carolina

Washington University School of Medicine
St. Louis, Missouri

Appendix B

EVALUATION MEASURES

1. Introduction

This section describes an approach to identifying relevant indicators of scientific productivity for evaluation purposes. A model of the research cycle is presented. The remainder of the section describes operational approaches to obtaining these indicators and producing variables for the evaluation analysis.

1.1. Research Cycle Concepts

The utility of any NIH funding program must be evaluated in terms of its contribution to biomedical science. Productivity for individual investigators has long been measured by choosing relevant and available indicators of productive activities in the cycle of research: obtaining research funding, doing the research, disseminating the results, relying on the merit of the work to secure continued funding, and repeating the cycle.

The principal outcome measure reflecting dissemination of research results is the number of publications. Scientific worth or quality is most commonly measured by bibliometric analysis, usually citation impact. When evaluations have tracked the ability of researchers to use their results to secure additional research funding, the measure has been the number and dollar amount of awards, especially investigator-initiated research (R01) awards. Research performance, and how the research is done, is seldom measured although a few evaluations have measured collaborative publication. The present evaluation hopes to go beyond the usual measures such as bibliometrics wherever this is feasible.

1.2. Indicators

Operationalization of the variables is described the sections that follow.

2. Research Support

Research support refers to the number, type, and amount of awards for purposes of supporting research. Support for research comes in many varieties—research grants, training grants, career development grants, and construction and resource grants. For purposes of measuring the productivity and worth of research, R01 awards are considered the most prestigious and indicative of innovative and potentially influential science.

2.1. Data sources

Input for the process will be the names of potential principal investigators. These will be derived from three sources. For the ARCs, a list of key personnel will be recorded from the applications submitted by the ARCs. This list will be supplied to the ARCs and any additions or corrections will be requested. For the control group, clusters of NIAAA R01 recipients will be identified.

Awards for all ARC and control group key personnel will be searched in NIH's Information for Management, Planning, Analysis and Coordination (IMPAC) system. A parallel search may also be performed in the NSF Awards Database, if NSF allows it.

2.2. Measurement Operations

Input Data (Names of Investigators)

Alcohol Research Centers

1. Obtain lists of key investigators from all new and competing continuation awards for ARCs. Record name and institution in a database. (Source: NIH Center for Research Review or NIAAA).
2. Send list to ARCs requesting verification. Make changes to database as needed. (Source: ARCs)

Control Group

3. Use results of control group identification procedures noted above. Include names in database with ARC key personnel.

Output Data (Awards)

1. Access IMPAC. Search records for all named investigators for designated periods of performance.

2. Download information on Activity Code, Administering Organization, Serial Number, Year of Awards, Period of Performance, Principal Investigator, Project Title, Award Amount, Recipient Institution, and School/Department.

3. Calculate derived variables: Total Award Amount, Award Amount per Year, Award Amount per Investigator, Awards from NIAAA versus Other, NIH versus Other, R01 Awards, Research Project Awards, Other Awards (by Activity Code).

3. Research Performance

Research performance may refer to a variety of concepts related to carrying out research. Some related to the concept of the ARC include use of core resources, collaboration among investigators, and technology transfer (meaning transfer of core resources and other scientific technologies to investigators in other institutions).

3.3. Data sources

Input for the core resources and technology transfer is the list of core resources and technologies found in the applications and confirmed by a survey of programs. Input for the collaboration analysis is the list of key personnel from ARCs and the control institutions.

The source of data will be a survey of programs requesting the counts of use of core resources and technologies, technology transfer requests, and collaborations among investigators. Collaborations will be independently confirmed in review of publications.

3.4. Measurement Operations

Core Resources and Technology Transfer

Input Data (Core Resources)

Input is the list of core resources and technologies found in the ARC applications and confirmed by a survey of programs for each ARC and control institution.

Output Data (Core Resources and Technology Transfer)

1. Submit list of core resources and technologies to ARCs and control institutions.

2. Request records of use of core resources and requests for technology transfer in each program.

3. Calculate derived variables: Count of Uses of Core Resources, Average Use of Core Resources per Year, Number of Technology Transfer Requests.

Collaboration

Input Data (Collaboration)

Input for collaboration analysis is the list of key personnel from ARCs and the control institutions.

Output Data (Collaboration)

1. Submit list of investigators to ARCs and control institutions.

2. Request list of major collaborative projects and collaborative publications among key investigators. (Publications will be verified against bibliometric analysis.)

3. Calculate derived variables: Count of Collaborative Projects, Proportion of Investigators Involved in Collaborative Projects, Count of Collaborative Papers, Average Number of Collaborative Papers per Year, Average Number of Collaborative Papers per Investigator.

4. Dissemination of Results

Insofar as science produces new knowledge stored in verifiable form for later use by science and by society,[10] then the number of publications in refereed journals remains the fundamental unit of productivity measurement in science and has been used in numerous previous NIH evaluations. Not only is the count of publications a quantitative indicator of productivity, but also it forms the basis of quality measures such as citation count and impact and journal impact.

Dissemination also refers to knowledge transfer more generally. An important mechanism for knowledge transfer, and one that is formally supported by NIH, is the training of young scientists. The number of trainees is an important quantitative measure of dissemination.

4.5. Data sources

Publications

Input for the publications analysis is the list of key personnel from ARCs and the control institutions who are potential authors. The Institute for Scientific Information (ISI®) will search for publications in its database.

Training

Input is the list of ARCs and control institutions and results of searching the NIH IMPAC II database for training awards to the institution.

The source of data will be a survey of programs requesting the number of trainees associated with each program and the destination of postgraduate trainees after they have left the institution.

10 Perkowitz S. "Generating science: Productivity and policy," *The Scientist*, 7:11, 1993.

4.6. Measurement Operations

Publications

Input Data (Names of Investigators)

Input for the publications analysis is the list of key personnel from ARCs and the control institutions.

Output Data (Publications)

1. Submit list of names of investigators to ISI. Search records for all named investigators for designated period.

2. Download information on Author(s), Author E-mail Address (to verify identity if needed), Title, Source, Language, Publication Type, Current Contents Subset and Categories, Entry Week, Reprint Author, and Institution.

3. Calculate derived variables: Number of Publications, Average Publications per Year, Average Publications per Investigator, Number of Publications per Journal.

Training

Input Data (Names of NIH Training Programs)

Input is the list of NIH-funded training grants and their period of performance for each ARC and control institution.

Output Data (Trainees)

4. Submit list of training grants to ARCs and control institutions. Search records for all named investigators for designated period.

5. Request number of trainees trained by each program and their destination when they left the institution.

6. Calculate derived variables: Number of Trainees, Average Trainees per Year, Proportion of Trainees Going to Institution with an ARC versus Control Institutions versus Other.

5. Scientific Worth

Scientific worth refers to the value and importance of scientific findings. Scientific peers usually judge worth. One indicator of worth is the utilization of research results and measured by citation of published papers. Other measures include honors and awards made by professional organizations, the prestige of programs or institutions engaged in alcohol research as judged by investigators in the field of alcohol research, and the issuance of patents for research products.

5.7. Data Sources

Bibliometric Measures

Input for bibliometric analysis is the list of publications of investigators in ARCs and control institutions and the list of journals in which they are published. ISI will search for publications in the *Science Citation Index* (*SCI*®). This database provides access to current and retrospective bibliographic information; author abstracts, and cited references found in 3,700 of the world's leading scholarly science and technical journals covering more than 100 disciplines.

An alternate approach will seek nominations from alcohol researchers of the most important publications of the past decade in alcohol. Input for this analysis is a random sample of alcohol researchers (from principal investigators funded by NIAAA).

Peer Judgments of Programs

This analysis will seek nominations from alcohol researchers of the most important programs (institutions) in alcohol research over the past decade. Input for this analysis is the random sample of alcohol researchers (from principal investigators funded by NIAAA) noted above.

Honors/Awards

This analysis will assess significant honors and awards bestowed on alcohol researchers. It will seek nominations from alcohol researchers of the most prestigious honors and awards for alcohol research and research in their specific discipline. Sources of winners of the awards and honors over the past decade will be sought from the professional organizations that give the awards.

Patents

This analysis will identify any patents issued to the ARC institutions or control institutions. The input information will be the names of the institutions. The source of information will be PATSEARCH, an online database supported by the MicroPatent Company. This database contains all utility patents, reissue patents, and defensive publications issued by the U.S. Patent and Trademark Office since 1975.

5.8. Measurement Operations

Bibliometric Measures

Input Data-Citation Analysis (Names of Investigators)

Input for the bibliometric analysis is the list of key personnel from ARCs and the control institutions.

Output Data (Citation Analysis)

1. Submit list of key personnel from ARCs and control institutions. Search records for all named investigators for designated period.

2. Download information on citations and journal impact.

3. Calculate derived variables: Total Citations, Average Citations per Year, Average Citations per Investigator, Citation Impact (a standardized measure that counts citations in the current year for a standard prior period, usually two years), Average Citation Impact per Year, Average Citation Impact per Investigator, Average Citation Impact of Journals.

Input Data-Peer Nomination of Important Papers

Input for the peer nomination of important papers in alcohol research is a random sample of alcohol investigators and a questionnaire.

Output Data (Nominated Papers)

1. Draw a random sample of alcohol investigators (providing ± 5% precision and 95% confidence) from IMPAC II.

2. Send questionnaire to investigators asking for nominations for important papers of the last decade.

3. Calculate derived variables for each ARC or control institution: Total Nominations, Average Nominations per Investigator, Citation Impact, and Average Citation Impact of Journals.

Peer Judgments of Programs

Input Data (Peer Judgments)

Input for peer judgments of the top research programs and institutions in alcohol research is a random sample of alcohol investigators and a questionnaire.

Output Data (Peer Judgments)

1. Draw a random sample of alcohol investigators (providing ± 5% precision and 95% confidence) from IMPAC II.

2. Send questionnaire to investigators asking for nominations of the top programs and institution in alcohol research over the past decade.

3. Calculate derived variables for each ARC or control institution: Total Nominations, Ranking.

Honors/Awards

Input Data (Honors/Awards)

Input for honors and awards is the list of honors and awards provided by a random sample of alcohol investigators.

Output Data (Honors/Awards)

1. Submit list of key personnel from ARCs and control institutions. Search records for all named investigators and their institutions.

2. Download information on patents and date of issue.

3. Calculate derived variables: Total Honors/Awards, Honors/Awards per Year.

Technology Transfer/Patents

Input Data (Patents)

Input for the technology transfer analysis is the list of key personnel from ARCs and the control institutions and the names of the institutions themselves.

Output Data (Patents)

1. Obtain lists of institutional research products such as diagnostic tests, animal models, patents, etc. identified by principal investigators and other investigators.

2. Determine to the extent possible the actual frequency and scope of dissemination of these products.

3. Calculate derived variables each ARC or control institution: Total technology transfer products, total patents.

Appendix C

DATA COLLECTION INSTRUMENTS

Alcohol Research Evaluation — Investigator Interview

Hello, this is _____, calling from The Gallup Organization, on behalf of the National Institute for Alcoholism and Alcohol Abuse. May I please speak with **(name of Investigator from file)**?

(When qualified respondent is reached, continue:) NIAAA is interested in evaluating its funding programs. They have asked us to conduct a brief survey of investigators who are supported by NIAAA funds get feedback on how they can improve their programs.

This interview will be recorded for my supervisor to review the accuracy of my work.
Your answers will be kept completely confidential, and only qualitative generalizations or statistical totals will be given to NIAAA. May I ask you the questions now?

1. What kinds of support does (institution **S1 from file**) provide for alcohol research here?

2. To what degree is alcohol research a central part of the strategic plan for research at (institution **S1 from file**)?

3. What is the advantage to your research of being at (institution **S1 from file**)?

3a. What is the advantage, if any, of having other alcohol researchers at the same institution?

4. (ARC only) What is the advantage to your research of participating in an Alcohol Research Center?

5. What core resources are available for alcohol research at (institution **S1 from file**)?

5a. Have you yourself used these resources?

6. Have you developed any research tools, technology, or other material of use to other investigators? What were these?

6a. What other investigators have used them?

7. Have you won any awards or honors at (institution **S1 from file**) or from other organizations recognizing your research in the alcohol field?

8. Does (institution **S1 from file**) provide administrative support for your research?

9. Does (institution **S1 from file**) make available funds for pilot studies or other research development activities?

10. Do you have research support other than from NIAAA? Where is that support from?

11. Please tell me how, in your alcohol research, you collaborate with other investigators?

11a. Would you say that collaboration is either encouraged or discouraged at (institution **S1 from file**)? In what ways?

12. Describe the ways in which you have mentored or developed younger alcohol researchers, if any?

13. Describe the ways in which you have provided outreach to the academic or public communities around issues related to alcoholism and alcohol abuse.

13a. Was this related to your research in any way?

Alcohol Research Evaluation — Administrative Questionnaire

The Gallup Organization is assisting the National Institute on Alcohol Abuse and Alcoholism in evaluating its funding programs. They have asked us to conduct a brief survey of institutions with significant levels of NIAAA funding to get feedback on how they can improve their programs. We are sending this questionnaire to designated contacts in NIAAA-funded institutions to better understand about training, core resources, and technology transfer.

Please complete this questionnaire and return it to _____ as soon as possible. Fax your response to Dr. _____ at ___-___-____. If you have any questions, please call Dr. _____ at ___-___-____.

We would like to know the number of individual who have been trained in alcohol research at this institution. For each level of trainee—undergraduate, graduate, or postdoctoral--please indicate in the grid below the dates and number of trainees. Include trainees who were present for part of a year. Please use a consistent definition of year related to the academic schedule, grant administration cycle, or calendar year.

For example:

Beginning (month/yr)	End (month/yr)	Number of Trainees		
		Undergraduate	Graduate	Postdoctoral
9/03	Present	0	7	3
9/02	8/03	0	6	3

Please start with the most recent period and complete the grid for as many years as possible.

Beginning (month/yr)	End (month/yr)	Number of Trainees		
		Undergraduate	Graduate	Postdoctoral

We would like to know the placement or destination of postdoctoral fellows who have been trained in alcohol research at your institution. In the grid below, please list each

postdoctoral fellow, his or her dates of matriculation, and their immediate appointment after leaving your institution.

Fellow's Name	*Dates of Matriculation*		Field of Study	Destination	
	Beginning (month/yr)	End (month/yr)		Institution	Title
Ex. Joe Doe	8/01	7/03	Liver disease	Big City U	Asst Prof

We also would like to know about the development of new investigators in alcohol research (beyond the level of postdoc) in your institution in the past ten years. In the grid below, please list each faculty member of other investigator developed from new investigator to established investigator in alcohol research, his or her starting date as a new investigator, and field of study. If there was a particular person who mentored this individual formally or informally, please list them in the grid as well.

Investigator's Name	Date Started as New Investigator (month/yr)	Field of Study	Mentor (if any)
Ex. Joe Doe	8/98	Liver disease	Dr. Jones

Now we would like to know about the utilization of core resources related to alcohol research that have been developed and/or are administered in your institution. For each core resource, please describe what it is. Also, identify the investigators and projects in your institution and elsewhere, if applicable, who have used this resource. If frequency of use is relevant, please indicate that as well.

Copy this page and use a new page for each core resource.

Name of Core Resource:	

Description of Core Resource (including personnel responsible for the resource):

Core Resource User	Project Using Core Resource	Dates of Use	Amount or Frequency of Use
Ex: Joe Doe	Trial of Alcohol Antagonist	4/00-3/01	64 animals

We would like to know about the utilization of any other research tools, technology or

materials—such as diagnostic tests, animal models, questionnaires, or patents—developed in the course of alcohol research at your institution and how they have been distributed to the research community. For each such research product, please describe what it is. Also, identify the investigators and projects in your institution and elsewhere, if applicable, who have used this resource. If distribution has been widespread through sales or publication please describe. If frequency of use is relevant, please indicate that as well.

Copy this page and use a new page for each tool, technology or material.

Name of Product:	

Description of Product (including personnel and project responsible for the development of the product):

Core Resource User	Project Using Core Resource	Dates of Use	Amount or Frequency of Use
Ex: Joe Doe	Trial of Alcohol Antagonist	4/00-3/01	64 animals

Finally, we would like to know about any other ways in which alcohol research at your insitution has impacted the alcohol research community. If there are ways not alr3ady descorbed previously, please describe them here.

THANK YOU. PLEASE SEND THIS QUESTIONNAIRE TO ...

Appendix D

FEEDBACK FROM ALCOHOL RESEARCH CENTERS CONCERNING EVALUATION DESIGN AND FEASIBILITY

1. Introduction

1.1 Sources of Feedback

This section describes feedback on the Alcohol Research Centers Program Draft Evaluation Design, results of the Feasibility Study and various other feedback and comments received from consultant Dr. Richard Longabaugh, NIAAA staff and staff of other NIH institutes.

1.2 Summary of Feedback

Most respondents indicated that the information requested in the evaluation is already submitted to NIAAA in the ARC renewal applications. They did not believe that they should be asked to resubmit the same information, but that NIAAA should use the information for both purposes. It was suggested that NIAAA might restructure the renewal application to fit both purposes. Respondents indicated that they would prefer to have questions added to the renewal application than to be bothered with a redundant effort.

Several respondents expressed concern about the focus of the evaluation. They were not convinced that the parameters were well-defined. A recurring concern was that evaluations would not take the uniqueness of ARC research into consideration. They feel it is important to maintain an accurate reflection of focus of program.

All respondents expressed concern about the validity of the proposed control groups, noting explicit differences between ARCs and R01s. Many suggested that the selection of valid control groups would be nearly impossible. Two disparities between ARCs and clusters of R01s that were frequently mentioned are the budget size and the research motivation ("Centerness" / community atmosphere / outreach efforts). Several respondents indicated that it would be very difficult to identify significant clusters of R01s with comparable research focus.

Many respondents questioned the probability of accurately gauging productivity, citing disparities in research focus.

Respondents recommended organizing comparisons by:

- Discipline
- Budget size
- Seniority of investigators

- Productivity
- Type of research (basic versus applied)

2. Interviews with ARC Principal Investigators

2.1 Alcohol Research Center #1

Favorable evaluations could lead to external criticism
Won't all ARCs look the same?
Many elements are difficult to quantify because of capricious funding – the value-added by ARCs is qualitative.
There are limitations to comparisons – they are not random.
Comparisons of ARCs to R01 holders are not valid on any measures central to ARCs
There is a different pull between ARCs and group of R01s – awards/investigators

The fostering of young investigators / assistant professors should be considered in an evaluation

Comparisons should match the level of seniority of investigators
Compare aggregate productivity/impact of 15 researchers at the ARC to 15 individuals.
Match 1 in ARC to 1 with R01
Match junior faculty to junior faculty

Time elapsed as a mitigating factor– all research operations look better given more time because investigators tend to have more impressive credentials farther in their careers.

It might be useful to solicit input from outside – those who lost ARC application / those who don't like ARCs
A third comparison group might be those who successfully applied at one time and chose not to continue.

2.3 Alcohol Research Center #2

What is missing is an evaluation of "Centerness"
What unique contributions can ARC make?
How is uniqueness evaluated?
An evaluation is useless if it does not take uniqueness into account.
What can be achieved at ARC vs. outside of ARC?
Must do more than R01 to achieve unique objectives
Describe the value added by the Center

Important standards include:
Promotional collaboration
Stimulating new research / development of new programs / success of pilot projects

How effectively pilot projects are converted
Effectiveness of public outreach / public awareness
Addressing regional issues – for instance USC studies alcohol among Hispanic population

Categorize by disciplines – behavior science / alcoholism / minority focus

Budget size is an inherent difference between ARC and R01 ($100,000 vs. $200,000-$250,000)
How can productivity be accurately assessed with such a big budget discrepancy?
Standardize evaluation by the budget size?
How much additional funding is generated?

Strike a balance between qualitative measures and quantitative measures.
Get data for quantitative measures from renewal applications – add questions to applications if necessary, but don't bother with redundant effort.
Share data and set well-defined parameters.

2.3 Alcohol Research Center #3

The comparison groups are imperfect.
The selection of valid control groups would be nearly impossible.
It would be very difficult to find 15-30 comparison groups.
Centers have more leverage than R01s in negotiations.

Certain issues must be identified and dealt with qualitatively. For example, only one paper published, but it's seminal.

Productivity levels vary dramatically.
Measuring productivity will be very tricky.
Some activities commonly done by ARCs are not done by R01s, including outreach programs, such as dissemination information at meetings/symposiums/conferences.

The purpose of ARCS and clusters of R01s vary.
It is important to maintain an accurate reflection of focus of program - LSU does basic primary research – biomedical.

Consider putting together an advisory panel – look at data – sort out issues

2.4 Alcohol Research Center #4

Worried about the validity of control groups.
Consider using unsuccessful ARCs
Compare to clusters of R01s / trans-institutional clusters

Isolated R01 may not be relevant to compare because the rationale for an ARC is clustering of activities.
There are massive budget disparities between Center-supported institution / R01s clustered.
Part of the success of an ARC is in stimulating R01s, generating spin-off projects.
R01 funding increases by selecting groups with increasing trajectory – historical.

Controls are defined too broadly – they refer to entire universe. Must restrict based on unique focus.

Washington University works with University of Missouri – Columbia. They are the only PSO from other NIH institute with a P30.

It is important to recognize different types of research (basic versus clinical).

2.5 Alcohol Research Center #5

Design comparison of ARCs with RO1 awardees is more important than comparison with unsuccessful ARC applicants.

Need to covary for the amount of the award. Each component project is equivalent to an R01, might expect productivity to be equivalent to $200K – $250K R01.
Do not focus too much on principal investigator, since at this institution the PI is a university official who is not an alcohol researcher.

Equivalence with R01 awards may depend on the nature of the RFA
R01s can be risky, especially when responding to an RFA requesting new research
A Center is equivalent to responding to 4-5 R01s for one deadline; it is usually a one-shot deal because it is too much effort to resubmit.
Should compare to a cross-section of R01s—a random group of comparisons to represent both successful and unsuccessful R01s.

To what extent is the product the reporting of data—what about chapter or books? What about conference presentations?
Which papers are attributable to the Center and which are not?

Ambivalent about using funding from other awards as a measure of productivity. Worry that generating grants may be to the detriment of scientific productivity. Too many grants may be difficult to manage.

Citation is an objective measure of impact on the field. Reflects the broadest definition of peers—investigators, graduate students and practitioners.

Pilot project success is measured in degree to which funded NIH projects result. Pilot components generally treated harshly in ARC review process. Their Arc supported ad hoc pilot projects, reviewed and funded internally, that ended up being more successful.

Most relevant training is transition from postdoc to principal investigator. They support many K-award proposals. ARCs have also encouraged new junior investigators to become established in the alcohol field.

Expectation for Centers (versus R01s) is of providing service for the common good (pilots, new investigators, training)—more general contribution to the field.

3. Feedback on the Feasibility Test

1) Awards
Data is available in ARC renewal application. Data goes back ten years in biosketch.
It is not difficult to collect data.
It would take about five hours.

2) Core Resources
This ARC's only core resources are animals.
The level of difficulty would depend on how far back data is needed. They could go about one year, but don't keep data much longer.
The total number of animals used can be found in the ARC renewal.
It would take a few hours to pull the information together.

3) Collaborative Projects
Data available on the publications list in the ARC renewal application.
It would take six to seven hours to put together. They have an appendix with reprints that goes back five years. They keep the renewal applications, but not the appendices.
It would be more difficult for an outsider to do.
Investigators could provide this information individually, but would be very irritated.
Some investigators have left the ARC.

4) Training
There are two sources for this information: training grants and the ARC renewal application. Both are imperfect. Some graduate students are not included in grant (about 75% included).
The information could be collected at the graduate office.
It would take about ten hours.

5) Technology Transfer
Their only technology transfer is animal models. They are only considered technology transfer if they are mailed out.
They are not patented.

They best way to collect the desired information would be to pay someone at the institution. It is too many hours to ask an ARC employee to do for free. Outsiders would take much more time.

Other concerns: include the validity of proposed control groups.
For new ARCs – measure productivity of researchers before working at an ARC and then after working at an ARC.

It is not possible to conduct a purely valid evaluation of ARCs.
All information requested in the feasibility questionnaire is already available in ARC renewal applications.

4. Comments on the Evaluation Protocol

4.1 Anonymous

Overall, I think Gallup did a good job of presenting a design for your research center program evaluation. The model of the research cycle, the chart of concepts, indicators, and variables, and the more detailed description of data sources are well organized and easy to follow.

The discussion on page 3 about the ARCs serving as "historical controls for themselves" is an interesting quasi-experimental approach. Looking back at my Cook and Campbell (1979) text, this approach is similar to the repeated measures or interrupted time series design. The "repeated measures" design recommends two or more pre-measures of the outcome variable to ensure that the "before" or baseline measure is not unique in some historical aspect (e.g., a well-known scientist either comes or goes during the pre-center grant time period). I noticed that Figure 2.1 called for only one "O Time 1" measure. Is there a good argument for only one pre-measure? If possible and practical, I would recommend additional pre-grant measurements. The section on Temporality in Gallup's concept paper on the centers evaluation discusses this point further.

The concept of a "cluster of NIAAA funded R01 awards" as a control group has a lot of potential for defining a control group of institutions that has not had the presumed benefits of center type grant components (i.e., administrative and science cores, research projects, and pilot projects).

On defining and choosing a control cluster, do you anticipate difficulties in selecting R01s for a cluster that is independent of an ARC? Is it not likely that the ARCs, on specific research topics, have principal scientists who are already collaborating with principal investigators or associates on R01s from NIAAA or elsewhere? Are Center scientists also possibly linked collegially to other independent R01 PIs who may be linked to the R01 PIs in the control group cluster? The ARCs and clusters of R01s could be defined as complex research networks. The ideal design is one where two networks (one with a center grant and the other without) are known to operate independently. In such a design, the interaction between scientists who have ongoing working relationships would not compromise the findings.

Would it be feasible to identify a group of R01s that function as a well-known informal

collegial network and that do not have any grant or contract type collaborations. This group would have all of the properties of a functioning research network but without any formal organizational properties typical of a research center. If this were possible, the centers evaluation design might go more in the direction of a "natural experiment" (I have reference on this if you wish).

The model on page 8 of the scientific research cycle is clear and concise. The link between "Scientific Worth" and "Research Support" is important and may have parallels with the concept of "research and human capital trajectories" that Barry Bozeman is working to apply to
NICHD's infertility research centers program evaluation. It will be interesting to see if (and how) his technique of "research value mapping" might contribute to measuring the relationship between scientific worth and research support in the Gallup model.

The discussion of data sources on grant awards (p. 10) and trainees (p. 12) mention the IMPAC database, but do not mention the Consolidated Grant Applicant File (CGAF) and the Trainee and Fellow File (TFF), maintained by the Office of Extramural Research (OER).

4.2 Anonymous

Introduction, page 2. The evaluation questions are not precise enough. Expand section and be more precise.

Page 3. Do we have any empirical evidence that five years is the appropriate length of time for selecting a research center to evaluate?

Page 4. Delete the last paragraph if we have decided that without the funding, it is not possible to compare these groups.

Page 4, Recruitment. What if the ARCs refused to participate? This is a question we should bring up at this weeks' meeting.

Page 5, 1.3.1. Selection Criteria. Select a group of RO1s that fall in the same science area. It may be important to match R01 and center investigators according to confounding variable such as the stage of the investigator's career, the academic qualifications, etc. You can't match on everything so select what are the important control factors. It might be good to test along the within group variation to get some idea of how confounded the results are.

1.3.2. Recruitment. We should precisely indicate whom we would contact for which questions. This paragraph indicates the design is not as precise as desired. It is not a good idea to let the organization select, as this could bias the results.

1.4 Data Sources. Investigator questionnaire. The concepts need more precision.

Peer Survey. The perceived quality and prestige is weak and likely to be driven by the parent institution's quality and prestige. Perhaps we should consider focus groups, expert panels, content analyses, or blind review of some outputs.

Administrative Questionnaire. What is the research question here?

Page 10, Table 3.1. Also, page 13. Collaboration. If collaboration is considered good, why? Measuring collaboration is not really meaningful in ___field because it happens close to 100% of the time. Does it matter whom the collaboration is with? What is the norm in the field?

Nominated Papers. Good idea.

Honors/Awards. What kind and given by whom?

Page. 12, Core Resources. What is this? Is more better?

2.5 Dissemination of Results. What is the research question the data is supposed to answer? How do you account for softer science approaches and alternative medicine approaches? You could look not only at the aggregate measures but also the same measures compared to the overall measures for all publications in that journal.

Trainees. What is a trainee? Formal pre and post doc training or does informal training count?

Publications, page 14. Are all publications counted or just original work? What about reviews, editorials, commentaries?

Page 14. IMPAC II. There are problems with the IMPAC II data sources for training information. Do not rely completely on this.

Page 17. Output data. There are implications for what is included and what is not.

What is NIAAA's minimal acceptable response rate?

Page 18. One cannot draw a sample of alcohol investigates from IMPAC II but can draw NIAAA investigators. What are the criteria? How big a sample and what is the response rate?

Page 20. Peer Review. What is the purpose of the site visits. For what purpose? What data will be gathered and how will it be used? What research a question is the data designed to address?

4.3 Anonymous

In forming the comparison group/R01 "cluster", how do you define "institution"? Broadly, so all parts of Johns Hopkins are the same "institution"? By major component (medical school, e.g.)? By department? There are pros and cons, of course, for each but it will make a major difference.

5. Feedback on Draft Paper on Alternative Evaluation Designs

5.1 Johnson, Paul

NICHD

1. Overall, the paper is an interesting exploration of possible designs to address the ARC evaluation questions. Some of the limitations of experimental/control designs mentioned in the paper are similar to the issues we are facing in the review of NICHD center programs.

2. The discussion about using "unsuccessful ARC applicants" as controls is worth further development. To me, the most challenging question is whether the two groups (successful and unsuccessful applicants) are sufficiently equivalent on some measure of potential for scientific productivity. On page 2, in the section on Comparability, I am not sure about the argument that unsuccessful applicants that met basic eligibility criteria for the programs are the same as successful applicants who met the same criteria. Being a successful applicant could be considered an important selection factor that interacts with the experimental treatment (i.e., the center grant).

3. The evaluation design that compares Centers versus RO1s is a difficult issue. One could argue that it is an "apples versus oranges" comparison. The approach worth exploring further, however, is the concept of a "cluster of RO1s" that could be defined as the equivalent of the "research component of a center." That is, the RO1 cluster could be considered as the same as the treatment group, but absent the other center-like components (administrative core, shared resources core, and the pilot project component, pp. 11-12).

4. There is only one mention about the "training of young scientists" (page 12). Is this an important outcome objective of the ARC program? Measures of this outcome were not discussed in the draft paper.

5. One of the ARC evaluation questions uses the expression, " value added of a having a center." Is Gallup charged with developing methods to measure and assess "value added"? The concept implies an examination of the cost efficiency of "centers" vs. other mechanisms.

6. The draft paper mentions that Gallup found about a dozen past center evaluations. Did Gallup provide you with a bibliography of those studies yet?

6. Comments on Miscellaneous Issues

IMPAC goes back about 5-8 years, after that the info would have to be gotten from Tina. To get a score, the PI would have to be called. We CAN give the names of the unfounded PI's, but scores are confidential. So we are left with depending on Tina for historical information past the last review five years ago. Hope this helps. We may have to rethink another comparison group, as several people have now expressed skepticism about looking at non-funded Centers.

We might want to have a meeting about comparison groups with Gallup and staff here, as Tina says Centers which have not been funded are not good comparison groups because they are all over the map...different themes, etc.
She suggested comparing to other Centers in other Institutes but we nixed that earlier. Also, some Centers do BETTER when not funded because it spurs them on to improve their Centers...

What if I just ran some really big whopping literature searches on the broad fields the centers are in (oral, head and neck cancer, e.g.), excluded reviews, dumped the results (unfiltered) into a database, pulled out statistics about what journals showed up in the broad search and compared these journals and their associated citation figures with the same statistics on what the centers publish? Pluses are that I think it is doable (I may have to write a couple of subroutines, or pay a programmer, but that's not much) and it would give us SOME comparison point for SOMETHING. Negatives are it only gives you comparisons on what journals are used and what the citation rates are, but no content info; in addition, because I couldn't filter the big searches, some things will be included that would have been filtered. Thoughts?