

# **NIH External Constituency Surveys - OER**

**Mini Supporting Statement B  
UNDER GENERIC CLEARANCE 0925-0627  
NATIONAL INSTITUTES OF HEALTH**

Name: Dr. Paula Goodwin  
Address: Office of Extramural Research  
RKL1 - One Rockledge Center, 3532  
6705 Rockledge Dr  
Bethesda, MD  
Telephone: 301.496.9232  
Email: [paula.goodwin@nih.gov](mailto:paula.goodwin@nih.gov)

July 2015

# Table of Contents

Section	Page
B. COLLECTIONS OF INFORMATION EMPLOYING STATISTICAL METHODS.....	2
B.1 RESPONDENT UNIVERSE AND SAMPLING METHODS.....	2
B.1.1 Respondent Universe.....	2
B.1.2 Sample Selection.....	3
INITIAL SAMPLE SIZES BASED ON PRECISION REQUIREMENTS.....	4
B.1.3 Response Rates.....	10
B.1.4 Sample Weights.....	10
B.1.5 Estimation Procedure.....	10
B.2 PROCEDURES FOR THE COLLECTION OF INFORMATION.....	10
B.2.1 Data Collection Procedures.....	11
B.3 METHODS TO MAXIMIZE RESPONSE RATES AND DEAL WITH NON-RESPONSE.....	11
B.4 TEST OF PROCEDURES OR METHODS TO BE UNDERTAKEN.....	12
B.5 INDIVIDUALS CONSULTED ON STATISTICAL ASPECTS AND INDIVIDUALS COLLECTING AND/OR ANALYZING DATA.....	12

## **B. COLLECTIONS OF INFORMATION EMPLOYING STATISTICAL METHODS**

### **B.1 Respondent Universe and Sampling Methods**

#### **B.1.1 Respondent Universe**

There are three populations of interest under the Peer Review Enhancement Surveys: an applicant population, reviewer population and Advisory Council member populations. These populations are defined as follows:

##### **Applicant Population**

The applicant population comprises those individuals who submitted R01, R03, R21, U01 and R34 applications to NIH reviewed in any of the Advisory Councils/Boards of NIH's constituent Institutes and Centers (ICs) in January or May 2015.

##### **Reviewer Population**

The reviewer population comprises those individuals who served in NIH study sections that reviewed R01, R03, R21, U01 and R34 applications that were subsequently reviewed by the Advisory Councils/Boards in January or May 2015. The target population of reviewers includes regular (appointed/permanent) and *ad hoc* (temporary) reviewers.

##### **Advisory Council Member Population**

The Advisory Council member population includes members who are regular (chartered) members and served at one of the NIH National Advisory Council/Board meeting held in January or May 2015. All Advisory Council members will be invited to take the Advisory Council survey. We anticipate that the number of members who will be eligible to take the Advisory Council survey will be 250.

##### **Applicant and Reviewer Population**

There are some individuals who are eligible to be members of both the applicant population and in the reviewer population. The sampling design for the peer review surveys was developed so that no individual who resides in both populations would be contacted for both the Applicant Survey and the Reviewer Survey. Table B.1-1 shows the total number of individuals in the universe of all applicants and reviewers (column 2), the number of individuals who are applicants but not reviewers (column 3) and the number of individuals who are reviewers but not applicants (column 5). Table B.1-1 also shows the numbers of individuals by race and ethnicity<sup>1</sup> in the total applicant population (column 6), and in the total reviewer population (column 7).

It also is possible for an Advisory Council member to be an Applicant and in some rare cases a Reviewer. Any eligible Advisory Council members who appear in the Applicant and Reviewer sampling frame will be removed from the sampling frame prior to drawing the sample of

---

<sup>1</sup> 1,746 Individuals with Unknown Hispanicity were assumed to be non-Hispanic for purposes of sample selection. 7,523 Individuals with Unknown Race were included in the "Other" race category along with 29,504 Whites and four multi-racial individuals who were not classifiable as Asian, Black, Native American, or Pacific Islander. Four Hispanic, Pacific Islanders were classified as Hispanic, Other for purposes of sample selection because of the extremely limited number of individuals in this group.

Applicants and Reviewers who will be invited to participate in the Applicant Survey and the Reviewer Survey.

The total number of applicants and reviewers (36,426) is equal to the sum of the number who are applicants only (22,439), the number who are reviewers only (7,779), and the number who are both applicant and reviewer (6,208). The number who are applicants (28,647) equals the number who are applicants only (22,439) plus the number who are both applicants and reviewers (6,208). The number who are reviewers (13,987) equals the number who are reviewers only (7,779) plus the number who are both applicants and reviewers (6,208).

Table B.1-1. Applicant and Reviewer Population Counts

Col. (1)	(2)	(3)	(4)	(5)	(6)	(7)
Stratum	All Applicants and Reviewers	Applicants Only	Both Applicant and Reviewer	Reviewers Only	Total Applicant Population	Total Reviewer Population
African American, Hispanic	23	13	2	8	15	10
American Indian/ Alaska Native, Hispanic	19	12	5	2	17	7
Asian, Hispanic	30	20	7	3	27	10
Multiracial, Hispanic	50	30	8	12	38	20
Native Hawaiian/ Pacific Islander, Hispanic	3	2	0	1	2	1
Other, Hispanic	1,393	808	264	321	1,072	585
African American, non-Hispanic	722	414	104	204	518	308
American Indian/ Alaska Native, non-Hispanic	45	24	8	13	32	21
Asian, non-Hispanic	7,078	4,572	1,303	1,203	5,875	2,506
Multiracial, non-Hispanic	304	181	60	63	241	123
Native Hawaiian/ Pacific Islander, non-Hispanic	30	23	3	4	26	7
Other, non-Hispanic	26,729	16,340	4,444	5,945	20,784	10,389

Total	36,426	22,439	6,208	7,779	28,647	13,987
-------	--------	--------	-------	-------	--------	--------

**Note:** A total of 7,942 persons with unknown ethnicity were assumed to be non-Hispanic for purposes of sample selection. 4,781 persons with unknown race were included in the “other” race category, together with 23,341 Whites.

## B.1.2 Sample Selection

### Determining Overall Sample Sizes

The total number of individuals who may be sampled and subsequently surveyed is defined by burden limits under NIH’s Office of Management and Budget (OMB) Generic Clearance No. 0925-0627. For the applicant and reviewer surveys, the total number of persons who may be sampled under the burden limits established by the NIH guidance is 4,460. Given the total number of allowable sample members, the next step is to determine how many of the allowable 4,460 should be selected from the applicant population and how many should be selected from the reviewer population.

### Broad Allocation Scheme

The total number of individuals to be sampled will be allocated to the following sets of individuals:

1. Applicants only (22,439)
2. Reviewers only (7,779)
3. Those who are both Applicant and Reviewer (6,208)

Within each set, sample sizes must be sufficient in order to allow for estimates within race and ethnicity groups to meet precision requirements (discussed below).

### Initial Sample Sizes Based on Precision Requirements

The following four steps will be taken for the three groups of sample members—applicants only, reviewers only, and those who are both reviewers and applicants:

1. A cross-tabulation will be created of the number of individuals by race (Asian, Black, Native American, Pacific Islander, “other,” and multiracial) and ethnicity (Hispanic or non-Hispanic).
2. Using the PracTools package (Valliant, Dever, Kreuter 2015) in the R computing environment (R Core Team 2014), the number of individuals required to be sampled in each race-by-ethnicity group will be estimated such that, within each group, a two-sided 95% confidence interval for a population proportion of 50% will have a half-width of 5%.
3. For those groups with population counts of less than 60 or for which the sample size is not large enough for the sample calculation to be approximately normally distributed, all group members in the relevant sample will be included. Such groups are said to be *selected with certainty*.
4. For those groups not selected with certainty, the PracTools package (Valliant, Dever, Kreuter 2015) in the R computing environment (R Core Team 2014) will report the required sample size.

The number of individuals selected with certainty or estimated by the PracTools package (Valliant, Dever, Kreuter 2015) in the R computing environment (R Core Team 2014) as being required in order to meet the precision requirement outlined in Step 2 are tabulated (Table B.1-2). These are the margin of error sample sizes.

**Table B.1-2. Initial Sample Sizes Based on Precision Requirements**

Stratum	Applicants Only		Applicants and Reviewers		Reviewers Only	
	Population Count (1)	Sample Size (2)	Population Count (3)	Sample Size (4)	Population Count (5)	Sample Size (6)
African American, Hispanic	13	13	2	2	8	8
American Indian/Alaska Native, Hispanic	12	12	5	5	2	2
Asian, Hispanic	20	20	7	7	3	3
Multiracial, Hispanic	30	30	8	8	12	12
Native Hawaiian/Pacific Islander, Hispanic	2	2	0	0	1	1
Other, Hispanic	808	261	264	264	321	176
African American, non-Hispanic	414	200	104	104	204	134
American Indian/Alaska Native, non-Hispanic	24	24	8	8	13	13
Asian, non-Hispanic	4,572	355	1,303	594	1,203	292
Multiracial, non-Hispanic	181	124	60	60	63	55
Native Hawaiian/Pacific Islander, non-Hispanic	23	23	3	3	4	4
Other, non-Hispanic	16,340	376	4,444	709 <sup>1</sup>	5,945	361
Total	22,439	1,440	6,208	1,764	7,779	1,061

The sample sizes listed in Column 4 will be subsampled such that half of the individuals selected will be assigned the applicant questionnaire; half, the reviewer questionnaire. The sample sizes listed in

Column 4 for those race-ethnicity groups not selected with certainty are designed so that the precision requirements will still be met for the 882 (half of 1,764) individuals selected to receive the applicant questionnaire and for the 882 others selected to receive the reviewer questionnaire.

Column 5 shows the number of individuals who are reviewers only, and Column 6 lists the numbers of individuals who must be sampled to achieve the stated precision requirement.

In addition, comparisons between groups will be required. We need about 400 (388 per group) applicants or reviewers per group to detect a 10 percentage point difference, i.e., group one equal 0.5 and group two equal 0.6, at power equal to 0.8 with alpha equal to 0.05 for a two-sided test. Consequently, the margin of error samples sizes that are not 400 will have to be increased to 400 to create the group comparison sample sizes. The group comparison sample sizes are shown in Table B.1-3, which has the same format as Table B.1-2.

**Table B.1-3. Group Comparison Sample Sizes**

Stratum	Applicants Only		Applicants and Reviewers		Reviewers Only	
	Population Count (1)	Sample Size (2)	Population Count (3)	Sample Size (4)	Population Count (5)	Sample Size (6)
African American, Hispanic	13	13	2	2	8	8
American Indian/Alaska Native, Hispanic	12	12	5	5	2	2
Asian, Hispanic	20	20	7	7	3	3
Multiracial, Hispanic	30	30	8	8	12	12
Native Hawaiian/Pacific Islander, Hispanic	2	2	0	0	1	1
Other, Hispanic	808	400	264	264	321	321
African American, non-Hispanic	414	414	104	104	204	204
American Indian/Alaska Native, non-Hispanic	24	24	8	8	13	13
Asian, non-Hispanic	4,572	400	1,303	800	1,203	400
Multiracial, non-Hispanic	181	124	60	60	63	55
Native Hawaiian/Pacific Islander, non-Hispanic	23	23	3	3	4	4

Other, non-Hispanic	16,340	400	4,444	800	5,945	400
Total	22,439	1,919	6,208	2,061	7,779	1,431

### Allocating the Remaining Sample of 400 by Consideration of Weighting

The remaining 400 individuals will be allocated to the three possible samples according to the impact of sampling weights on the precision of estimates generated from each of the samples. A widely used measure for assessing the degree to which sampling weights affect the precision of statistical estimates is known as the design effect. The design effect for a given statistical estimate is the ratio of the variance of the estimate under the appropriate complex sampling process to the variance of the estimate when it is assumed that the underlying data arose by means of a simple random sample. Our particular focus here is on the unequal weighting effect which uses the variability in the weights to assess the detrimental impact of the differential weighting. It is defined as follows: Given a sample of  $n$  individuals with a set of associated sampling weights denoted  $w_i$  and with the average sampling weight denoted, then

$$\text{Deff} = 1 + \left( \frac{1}{n} \right) \left( \frac{\sum_{i=1}^n (w_i - \bar{w})^2}{\bar{w}^2} \right)$$

In the case of simple random sampling, the design effect is 1. When design effects are larger than 1, then the variances of estimates are larger than they would be if the samples had been selected with simple random sampling. A general guideline is to try to keep the Deff from exceeding 1.50 (Valliant, Dever, and Kreuter 2013, page 375). In accordance with this consideration, additional individuals will be allocated to reduce this design effect.

With the group comparison sample sizes being calculated with use of PracTools package (Valliant, Dever, Kreuter 2015) in the R computing environment (R Core Team 2014), the design effect for the set of applicant-only individuals is 2.75, the design effect for the reviewer-only individuals is 2.12, and the design effect for those individuals who are both reviewers and applicants is 1.46. Because the design effect for those individuals who are both applicants and reviewers is the smallest and is already less than 1.50, additional sample will be allocated to the applicant only and reviewer only groups so that the design effect for three groups will about 1.46. By selecting additional applicant-only individuals from the “other,” non-Hispanic group to make the sample size 1,290 and additional reviewer-only individuals from the “other,” non-Hispanic group to make the sample size 810, the design effects of the applicant-only and reviewer only groups decreases to 1.46, which is similar to the group that contains both the applicants and reviewers. The “other,” non-Hispanic group’s sample proportions differ most from their corresponding population proportions, so allocating the extra sample to this group provides the greatest reduction to the design effect. The resulting unequal weighting effect sample sizes are summarized in Table B.1-4. These are the sample sizes that we expect to have for the analysis. The total analytic sample size, i.e., the unequal weighting effect sample size, sample size is 6,711 (2,809 + 2,061 + 1,841).

## Sample Power Analysis

Based on the unequal weighting effect sample sizes, i.e., before the nonresponse sample size inflation, some power estimates are reported to illustrate the power to detect differences between some race-ethnicity groups (Table B.1-4). Power estimates are calculated under the assumption of simple random sampling, the sample allocation, and an unequal weighting effect of 1.46. The pwr package (Champely 2015) in the R computing environment (R Core Team 2014) was used to estimate the power associated with a two independent proportions, assuming unequal observations per group. The power is calculated with a two-sided test, with an alpha level of 0.05 and under the assumption that one of the groups has an underlying true proportion of 0.5. Estimates of power to detect 5 percentage point differences assume that the second group in the test has a true underlying proportion of 0.55. Estimates of power to detect 10 percentage point differences assume that the second group in the test has a true underlying proportion of 0.60. For each pair of strata for which power is calculated, the pwr package (Champely 2015) in the R computing environment (R Core Team 2014) requires the input of sample sizes within each stratum. Note that these are the minimum power calculations because they do not include the finite population correction, which would decrease the standard errors and increase the power to detect differences.

Table B.1-4 (A) shows, for the applicant-only sample, estimates of the power to detect 5 and 10 percentage point differences in proportions of the pair of strata (excluding certainty strata because no uncertainty among groups exists where a census is taken).

**Table B.1-4. Power Estimates for Detecting Intergroup Differences**

Stratum	Power to Detect 5% Difference	Power to Detect 10% Difference
<b>A. Applicant-only sample</b>		
Asian, non-Hispanic ( <i>n</i> = 400)	42	94
Other, non-Hispanic ( <i>n</i> = 1,290)		
<b>B. Reviewer-only sample</b>		
Asian, non-Hispanic ( <i>n</i> = 400)	37	91
Other, non-Hispanic ( <i>n</i> = 810)		
<b>C. Applicant-and-reviewer sample</b>		
Asian, non-Hispanic ( <i>n</i> = 260)*	24	72
Other, non-Hispanic ( <i>n</i> = 400)		
<b>D. Sample of applicant-only members and applicant-and-reviewer members to receive applicant-only questionnaires</b>		
Asian, non-Hispanic ( <i>n</i> = 660)	59	99
Other, non-Hispanic ( <i>n</i> = 1,690)		
<b>E. Sample of reviewer-only members and applicant-and-reviewer members to receive reviewer-only questionnaires</b>		
Asian, non-Hispanic ( <i>n</i> = 660)	54	99
Other, non-Hispanic ( <i>n</i> = 1,210)		

\* Certainty sampling stratum.

Similarly, Table B.1-4 (B) shows, for the reviewer-only sample, estimates of the power to detect 5 and 10 percentage point differences in proportions of the two non-certainty strata in the applicant-reviewer sample. Only the Asian, non-Hispanic group and the “other,” non-Hispanic group were not selected with certainty (see Table B.1-2). Power calculations are presented only for purposes of comparing these two groups, because no uncertainty is associated with estimation for groups for which a census is taken.

Table B.1-4 (C) shows, for the applicant-reviewer sample, estimates of the power to detect 5 and 10 percentage point differences in proportions of the two non-certainty strata in the applicant-reviewer sample. Only the Asian, non-Hispanic group and the “other,” non-Hispanic group were not selected with certainty (see Table B.1-2). Power calculations are presented only for purposes of comparing these two groups, because no uncertainty is associated with estimation for groups for which a census is taken.

Table B.1-4 (D) addresses the aggregation of sample members who are applicants only with those sample members who are applicant-reviewers selected to receive the applicant questionnaire; it shows estimates of the power to detect 5 and 10 percentage point differences in proportions of the pair of strata with the largest sample sizes (excluding certainty strata because no uncertainty exists among groups for which a census is taken). Because only two strata in the sample of individuals who are applicant-reviewers are not selected with certainty, power calculations are shown for only those two strata.

Table B.1-4 (E) addresses the aggregation of sample members who are reviewers only with those sample members who are applicant-reviewers selected to receive the reviewer questionnaire; it shows estimates of the power to detect 5 and 10 percentage point differences in proportions of the pair of strata with the largest sample sizes (excluding certainty strata because no uncertainty exists among groups for which a census is taken). Because only two strata in the sample of individuals who are applicant-reviewers are not selected with certainty, power calculations are shown only for those two strata.

### B.1.3 Response Rates

The response rates for the survey will be calculated in accordance with the recommendations that the American Association for Public Opinion Research (AAPOR) has published in its *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys* (2008). The formula for the response rate is as follows:

$$RR6 = \frac{(I + P)}{[(I + P) + (R + NC + O)]},$$

where  $I$  = complete interview,  $P$  = partial interview,  $R$  = refusal,  $NC$  = noncontact, and  $O$  = other nonresponse. Notably, this formula differs from the AAPOR formula RR4 in that, because all individuals in the NIH-provided sampling frame are assumed to be eligible for the study, no estimate of the number of eligible individuals among those with unknown eligibility is included in the denominator. Adjustments to the response rate formula can be made if ineligibility of some individuals is later determined.

### B.1.4 Sample Weights

Discussed here is the method to be followed to create the final sample weights and final estimates for the peer review surveys. One nonresponse-adjusted sample weight will be created for the applicant sample; another weight, for the reviewer sample. These weights will consist of a product of two factors: the base weight and the nonresponse adjustment, defined as follows:

- The *base weight* (for a given sample) is the inverse of the unconditional probability of selecting a sample member into the sample. This weight accounts for the stratification used in the sample design. Notably, if all sampled individuals respond, then no nonresponse adjustment is necessary.
- The *nonresponse adjustment* (for a given sample) is an adjustment imposed on the sampling weight of the respondents to account for those applicants who do not respond to the survey. In general, this adjustment will be greater than 1 so that each respondent will represent himself or herself, as well as some portion of the nonrespondents.

There are numerous ways of constructing a nonresponse adjustment. For each of the applicant and reviewer samples, the plan is to adjust the base weights within strata and to use a simple ratio adjustment. In order to perform this adjustment, we will need to know which stratum each respondent belongs to.

### B.1.5 Estimation Procedure

After the data are collected, analysis of the data must rely on software that can account for the sample design. Data analysis will be performed with SUDAAN software (2012). SUDAAN can manage

correlated observations in a general sense, with nonparametric and parametric approaches being available. Base SAS software (SAS Institute 2012) will be used for data manipulation and tabulation of results.

## **B.2 Procedures for the Collection of Information**

### **B.2.1 Data Collection Procedures**

Sample members will be asked to complete the surveys online. The basic steps involved in the data collection process for all three surveys include:

- An e-mail invitation will be sent to all sample members (Attachment 5). The invitation will be signed by a senior NIH official. It will explain the purpose of the survey and how they were selected to participate. It will invite the sample member to participate in the survey and will provide a hyperlink to the survey Website.
- Ten days after the e-mail invitation, a reminder e-mail will be sent to all sample members who have not submitted their responses to the survey. The e-mail will encourage those who have not yet logged in to the Website to participate in the survey.
- Ten days after the first e-mail reminder, a second e-mail reminder will be sent to all non-respondents. The e-mail will reinforce the purpose and relevance of the survey.
- Ten days after the second e-mail reminder, a third e-mail reminder will be sent to all remaining non-respondents.
- In addition, if the response rate for either the Applicant or Reviewer survey in any of the strata indicated above does not achieve 50%, a final reminder letter will be mailed by express mail (USPS Priority mail) to non-responders within the stratum, along with a hardcopy version of the survey. The reminder letter will contain the same information as the original invitation. Enclosed with the letter will be a postage paid, business reply envelope for returning the completed questionnaire.

### **B.3 Methods to Maximize Response Rates and Deal with Non-response**

The ability to gain the cooperation of potential respondents is key to the success of these two surveys. Consistent with sound survey methodology, the design of the survey will include approaches to maximize response rates, while retaining the voluntary nature of the effort. We will use the following approaches to maximize response rates for the surveys:

- Participation will be made as easy and non-burdensome as possible by designing each questionnaire to take no more than an average of 30 minutes to complete.
- The online instruments will be designed to be clear and easy to understand. Thorough usability testing of the survey instruments will be conducted to eliminate technical errors and to ensure ease of navigation and use.
- Advanced outreach will raise awareness about the surveys and to encourage participation (e.g., announcements on NIH Websites and newsletters and the OER RockTalk blog).
- The introductory e-mail invitations will inform sample members of the study. They will contain enough information to generate interest in the surveys. The letter and email will provide a point of contact at RTI for additional information.

- Follow-up e-mails will remind sample members about the survey, and encourage participation. These reminders will always include a link to the survey.
- A final reminder letter, if needed, will include a hardcopy version of the survey to provide an alternative mode for answering the questions.

#### **B.4 Test of Procedures or Methods to be Undertaken**

The survey instruments have also been tested through a modified Question Appraisal System (QAS). With the QAS, the questions in the instrument were analyzed in relation to the tasks required of the respondents (to understand and respond to the questions) and evaluate the structure and effectiveness of the questionnaire form itself. RTI International's Question Appraisal System (QAS-04) was used to guide this instrument review. This coding system constitutes an item taxonomy that describes the cognitive demands of the questionnaire and documents the question features that are likely to lead to response error. These potential errors include comprehension, task definition, information retrieval, judgment, and response generation. This appraisal analysis was used to identify possible revisions in item wording, response wording, questionnaire formats, and question ordering/instrument flow.

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

**Darryl V. Creel**  
**Jennifer Cooney**  
RTI International  
3040 Cornwallis Road  
Research Triangle Park, NC 27709