

Supporting Statement Outline – Sample

**NOTE: Complete Part B for Survey ICR Requests**

**SUPPORTING STATEMENT – PART B**

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

If the collection of information employs statistical methods, it should be indicated in Item 17 of OMB Form 83-I and the following information should be provided in this Supporting Statement:

1. Description of the Activity

The potential respondent universe breaks down into three major groups. The first group includes employees of major private sector software corporations. We define this group as anyone who works for specific companies such as Facebook, Amazon, or Microsoft (often referred to as the ‘FAANG’ corporations). The second group includes employees of defense contractors and aerospace firms. We define this group as anyone who works for a private sector corporation that derives at least a quarter of its profit from defense contracts. The final group consists of anyone with a software engineering background. We define this group as anyone who has an appropriate job title (“software engineer”, “software architect”, etc.) or has a college degree in Computer Science (CS). Table 1 indicates how many individuals are estimated to be in each of these three groups.

| <b>Group</b>                             | <b>Estimated Size</b> | <b>Respondents</b> |
|--|-----------------------|--------------------|
| All Software Engineers                   | 1,365,500             | 1,500              |
| FAANG Employees                          | 1,070,000             | 1,500              |
| Defense and Aerospace Industry Employees | 2,500,000             | 900                |

We will recruit and enroll a non-probability sample. We will use validated commercial and public sources to identify a demographically balanced sample from these populations. We plan to accomplish this by using a combination of targeted advertising from websites such as Facebook or LinkedIn that collect detailed demographic information about their user base, as well as information purchased from commercial data sources that offer verified information about corporations and their employees for purchase, such as Dunn & Bradstreet, or marketing lead generations, such as LinkedIn’s Lead Generation tool.

Recruitment will occur through three mechanisms. First, we will place targeted advertisements on sites such as LinkedIn that collect demographic information about their user community. The options provided by these sites to target advertising align directly with our survey participant criteria; employees of FAANG corporations can be identified by their self-identified employer, employees of the defense and aerospace industry can be identified by their employer, job title (‘aerospace engineer’), or whether they hold a security clearance; software engineers can be identified simply by their job title or by whether they have received a computer science degree.

Second, using the information obtained from the commercial data sources, we will email invitations to participate in the survey to individuals who meet the criteria of our various survey population groups. These criteria will be the same as described in the targeted advertisement outreach.

For the population groups in this study, we do not have prior experience to base an estimate of response rates on. We are assuming a response rate of 10% from these groups, which is standard for web surveys and comparable to rates reported by other studies conducting online surveys. However, this may be too pessimistic since the nature of the topic may entice individuals in these groups to want to participate in the survey.

## 2. Procedures for the Collection of Information

- a. Sample allocation, stratification: Our sample of software engineers will be representative of three non-overlapping groups (or, *strata*) consisting of those: (a) employed at leading software firms, (b) employed at defense industrial base and aerospace firms, and (c) employed outside of firms not included in (a) or (b). At the sample selection stage, we will allocate sample sizes explicitly to each of these groups such that the sample for each group is large enough to allow for estimation of independent, stratified estimates with adequate precision, and to enable comparisons and statistical testing of survey estimates for different strata. Lastly, the sample sizes for the focus groups will be much smaller as these constitute the qualitative portion of this study.
- b. Estimation: The analysis plan includes survey estimation of mean, proportion, and computation of associated standard errors, margin of error (MoE), and 95% confidence intervals for each of the survey items or constructs. We will compute the survey estimates mentioned above for the three groups of software engineers (leading software firms, defense industrial base or aerospace, other) in a stratified survey estimation analysis step. We will also conduct statistical testing of survey estimates across groups using an appropriate test (e.g. t-test, z-test, chi-squared test, or multiple regression) for the type of survey item.
- c. Degree of accuracy: The following is a summary of sample size and precision calculations conducted by our team, using different values of precision (or, margin of error). Our assumption is that a margin of error of +/- 5% is acceptable; also, calculations were conducted for binary variables, with an assumed proportion of .50 (50%) which yields conservative sample sizes. We found that a sample size of 100 completes, with an estimated proportion of .50 (50%), yields a margin of error of 5%, or a 95% confidence interval of (45%, 55%) around the survey estimate. Thus,  $n=100$  is the minimum sample size. Similarly, a sample size of 250 or 500 completes would yield a margin of error of 3.2% and 2.2%, respectively. Margin of error gets smaller and precision improves with bigger sample sizes; also, precision is worst at 50% and better when survey estimate is 40% or 30%. When we computed smallest detectable difference with two groups of sample size 100 each, that value is 19%. That is, if one group has an estimate of 50%, the second group would have to be 69% or greater (difference  $\geq 19\%$ ) to yield a statistically significant difference. Larger sample sizes in each group will allow us to detect smaller differences as significant.
- d. We do not anticipate any unusual problems requiring specialized sampling procedures.
- e. Respondent burden for each component of this one-time data collection is anticipated to be low.

## 3. Maximization of Response Rates, Non-Response, and Reliability

Survey of Software Engineers. The contractor will employ various methods to improve the response rate to the survey sent to software engineers. The survey sample will be identified shortly before the data collection begins to ensure that the sample information is current.

The primary mechanism we will use to reduce non-response bias is emphasizing the salience of the survey topic to this group. Artificial Intelligence (AI) is a topic of great interest to this community and the ongoing debate about what role ethics should play in the advancement of this technology continues to draw significant interest from software engineers. We will also promise confidentiality and ensure anonymity to our respondents. Additionally, experience has shown that limiting respondent burden reduces non-response. The survey is estimated to take software engineers no more than 15 minutes to complete. The survey will ask only about issues critical to the analysis and all information requested should be readily available to the respondent. Finally, we do not plan to offer incentives to survey participants as a mechanism to increase response rates simply because we are unable to offer a sufficient incentive amount to motivate members of this already highly compensated community.

The survey data collection will include follow-up with non-respondents to maximize response rates. For instance, the contractor will send prompting letters and/or emails to selected sampled individuals who have not responded to requests. As needed, additional replicates will be added to ensure that an acceptable number of responses is achieved. We will follow up by email a maximum of three times, for a total of 4 potential contacts (including the initial outreach email). Respondents can opt out of further contacts by contacting either an email address or phone number which will be provided as part of the outreach effort. Respondents will not receive further contacts if they choose to opt-out or after they have completed the survey.

As this is not a probability sample, our calculations will not require design or sampling weights. Instead, we will consider an analysis of non-response bias and post-stratification weighting, as appropriate. Note that for such analysis, we need a set of characteristics that are observed for the respondents, as well as the underlying population. Estimates of bias will involve comparing distributions of each characteristic for the respondents and the population. If the two distributions are different, we can use an iterative proportional fitting algorithm (such as a raking algorithm) to ensure that the two are made more similar. Finally, the raking algorithm would generate a set of non-response weights to apply to the respondents; the variance of these weights will convert to a design effect, which will effectively reduce the nominal sample size. The goal of this exercise is to reduce bias without over-inflating the variance of these adjustment or non-response weights.

Focus Groups. Focus group recruitment will be separate from the survey administration. The contractor will conduct focus groups at industry conferences and at universities with computer science departments. For focus groups convened at industry conferences, flyers with information for the focus group, including the date, time, location, will be provided to conference organizers for distribution with other conference materials. For focus groups convened with students, we will similarly send flyers to related listservs and social media sites for posting. For both types of focus groups, we may also print flyers with focus group details for posting or distribution. All materials will include the RAND team's contact information and will invite interested individuals to contact RAND. Additionally, we may use commercial focus group vendors to identify and recruit participants. These vendors have databases of individuals and will call people and screen them for eligibility and interest in the focus group. Once contact is made with potential focus group participants, a short screening script will be administered to determine eligibility and to ensure, to the extent possible, that participants represent a mix of demographic characteristics.

To maximize attendance at the focus groups, the contractor will also identify focus group candidates shortly before groups occur. Candidates will be invited to participate by email, mail, advertisement, and/or telephone and follow-up to individuals. Individuals who are eligible and agree to participate will receive a confirmation after screening and reminder shortly before the group occurs.

Focus groups will last 90 minutes and be held at a location convenient for participants. The methods proposed for this data collection should yield also acceptable response.

#### 4. Tests of Procedures

The survey instrument was cognitively tested with nine respondents and questions were revised based on findings from the cognitive interviews.

#### 5. Statistical Consultation and Information Analysis

This study is being conducted by the RAND Corporation, under contract to the Department of Defense (DoD). The RAND Principal Investigator (PI) is Mr. James Ryseff and Dr. Eric Landree is the Co-PI; they will oversee design and analysis of the data. Mr. Ryseff will lead the analysis of the qualitative data from the interviews and focus groups. Dr. Ghosh Dastidar consulted on the statistical aspects of the survey sampling and will direct the statistical analysis of the data. Mr. Noah Johnson will conduct the statistical analysis of the survey data.

Contact information:

Principal Investigator: James Ryseff  
Technical Analyst  
RAND Corporation  
1200 South Hayes Street  
Arlington, VA 22202  
Office: 703-413-1100 x5717  
Fax: 703-413-8111  
Email: jryseff@rand.org

Co-Principal Investigator: Dr. Eric William Landree  
Senior Engineer  
RAND Corporation  
1200 South Hayes Street  
Arlington, VA 22202  
Office: 703-413-1100 x5078  
Fax: 703-413-8111  
Email: landree@rand.org

Statistician: Dr. Madhumita (Bonnie) Ghosh Dastidar, Ph.D.  
Senior Statistician  
RAND Corporation  
1776 Main Street  
Santa Monica, CA 90407  
Office: 310-393-0411 x7418  
Fax: 310-393-4818  
Email: bonnieg@rand.org

Statistical Analysis: Noah Johnson  
Associate Policy Researcher  
RAND Corporation  
1776 Main Street  
Santa Monica, CA 90407  
Office: 310-393-0411 x6862  
Fax: 310-393-4818  
Email: [njohnson@rand.org](mailto:njohnson@rand.org)