

Effective Sampling for SCAMPs

**Ron Radice
Software Technology Transition
www.stt.com
2007 SEPG
Austin, Texas**

Why This Presentation?

While the SEI now requires “Full Disclosure” of sampling in the ADS, the SEI is not providing sufficient guidance for sampling design, especially for large size organizations.

The need for sampling design guidance must be properly addressed or the SCAMPI community will see inconsistent results.

Stimulate discussion and suggest sampling guidance.

- As a start, a four step approach is suggested to sampling for SCAMPIs.**

Suggestions are not posed as the answer, but are positioned to open further discussion toward a needed guidance for sampling design.

Appraisal Scope [SEI1]

“The organizational *scope of the appraisal must include sample projects and support groups that are representative* of the implemented processes and functional areas being investigated within the organizational unit and that operate within a coherent process context. *The rationale for selecting these sample projects and support groups as representative of the organizational unit must be documented.*”

“Organizational unit size (i.e., number of people and number of projects) and sizes of projects and support groups (i.e., number of people) in the organizational scope must be documented as well as the percentage ratio of these two measures:

- **Population %:** the number of people in the organizational scope divided by the number of people in the organizational unit (x100)
- **Project %:** the number of projects in the organizational scope divided by the number of projects in the organizational unit (x100)”

“Identify factors critical to the implementation of practices in projects.”

This is the basic sampling guidance of record!

Why Sample?

Sampling is performed when a target population is either too large or the cost to obtain information from the full population is too high.

We need to sample when doing appraisals because of:

- Time
- Cost
- Convenience / Accessibility

We thus trade off these practicality factors versus having the certainty that we would have by getting information from the entire population.

How much can be traded off without compromising integrity?
Shouldn't there be limits to interpretation; i.e., guidance?

When is enough enough?

SCAMPI suggests that a subset of an Organization Unit (OU) be sampled for an appraisal, but if it is an inappropriate subset without guided sampling, it may be misrepresenting the Level of the organization; e.g.,

- 3 projects where an organization has 300 projects is a 1% sample
 - Is this enough?
- 300 interviewees when an organization has 10,000 people is a 3% sample
 - Is this enough?

**But we don't want to over sample either!
So what can we do that is practical and effective
while maintaining integrity?**

Key Principles

Sampling should follow basic principles to ensure integrity:

- Random-ness
- Representative-ness
- Timing of sample selections
- Checking for non-bias

Sampling is a generalization of the population or Organization Unit (OU) from a subset of (Focused (F) & Non-focused (NF) projects documentation and interviewees (FARs)).

Key issue is one of confidence of the generalization:

- How certain can we be that the findings from the sample will hold true for the entire OU, if we did not study; i.e., interview, everyone?
- How typical or representative is the sample of the OU?

Random-ness

Using Simple Random Selection

Simple random selection exists when any project and any individual in a project in the OU has as likely an opportunity for being selected as any other that can be included in the sample; thus reducing the likelihood of bias.

- There are many ways to perform simple random selection.
[THO1]

If for *any* reason a selection has a higher probability of being made, then this is not a random selection, does not reflect the population in the OU, and will bias the results; e.g., a SCAMPI is biased when only selecting:

- Participants who have been with the OU longer and “can speak better” about the organization
- Projects that are “key” financially to the OU
- The “large” projects
- The “best” projects

Representative-ness

1 of 2

OUs are often not homogenous.

- **Various factors bespeak an OU's nature**

Critical factors of interest include: [SEI1]

- **Application domains (or lines of business)**
- **Geographical breadth**
- **Disciplines (e.g., systems engineering, software engineering, or hardware engineering)**
- **Effort types (e.g., development, maintenance, or services)**
- **Project types (e.g., legacy or new development)**
- **Customer types (e.g., commercial or government agency)**
- **Lifecycle models in use within the organization (e.g., spiral, evolutionary, waterfall, or incremental)**

I would add there are other factors to consider; e.g.,

- **size of the project**

Representative-ness

2 of 2

The selected subsets sampled across critical factors must be representative of the OU; i.e., the sampling technique did not inadvertently cause a demographic misrepresentation of the OU; e.g.,

- A line of business was not included
- An geographic location was excluded
- Some project types are not reflected in the samples
- Only projects of certain sizes were included

While the critical factor samples probably will not precisely match the percentages in the OU for the various critical factors, they should approximate the sample percentages within the OU.

**There may be overlap across critical factors,
but we still want representative-ness.**

Sample Timing

Samples are more appropriately selected as close as possible to the time of engagement to maintain integrity, allowing time for documentation review. A sample selected months before the appraisal may not represent the OU's practice performance as well as those samples selected when the appraisal is administered.

- OUs can be rather dynamic; a project may only exist for weeks or a few months.**

It should be assumed that any initially determined interviewees are only potential participants until the interviews are held.

Some participants, such as Organization Process Owners and some managers, are required to be identified early on due to their roles in the organization and it is not likely these will change.

The principle of Sample Timing is potentially compromised as time between selection and appraisal is increased!

Checking for Bias

While focusing on random-ness and representative-ness will help remove biases, the selected samples should be reviewed to ensure that no biases have been inadvertently introduced.

Why we want to avoid bias:

- Bias affects how far an average statistic lies from the parameter it is estimating; e.g., how different the sample is from the OU. Errors from chance (natural variation) will cancel each other out in the long run, those from bias will not.**

As with all statistics, we do not try for precision, but we do try for accuracy.

How accurate can we be with SCAMPs?

Traditional Sampling Parameters

- **Confidence Level**
- **Sampling Error**
- **Confidence Interval**
- **Population**

Confidence Level

The confidence level (CL) tells us how sure we can be with the result after sampling a population.

A 95% confidence level means we can be 95% certain; a 99% confidence level means we can be 99% certain.

- Population size and sample size affect CL

Most work performed with sampling uses the 95% confidence level.

Can we use 95% CL with SCAMPI?

Sampling Error

“In statistics, when analyzing collected data, the *samples* observed *differ in such things as means and standard deviations from the population from which the sample is taken. This is sampling error* and is controlled by ensuring that, as much as possible, the samples taken have no systematic characteristics and are a true random sample from all possible samples.” - [WIK1]

In any one sample, it is virtually certain there will be some sampling error and we can never be certain exactly how large this error is. [DEM1]

In general, the larger the sample size, the smaller the sampling error tends to be:

- We need to have a sample size large enough so that sampling error will tend to be reasonably small.
- If the sample size is too small, the results will tend to be too inaccurate to be of much use.
- There is also a point of diminishing returns beyond which increasing sample size provides little benefit. Once the sample size is large enough to produce a reasonable level of accuracy, making it larger simply wastes time and money.

Contributors to Sampling Error

Problems of sample measurement, leading to error, include:

- The sample design; e.g., how the projects were selected and how the individuals are selected within the projects
- The non-availability problem (do people who are available differ on what we are measuring from the people who are not available; e.g., on CMMI practice performance?)
- Question wording during interviews
 - Including question order
- Deliberate, or unconscious, lying or false reporting by interviewees
- SCAMPI rating rules as they relate to sampling integrity

In theory, with a sample of reasonable size, one can say; e.g., with 95% certainty (CL) that the results have a statistical validity of plus or minus X percentage points (where X is the Confidence Interval) of what the result would be if the entire population had been appraised.

Confidence Interval

The confidence interval (CI) is the plus-or-minus figure usually reported in newspaper or television opinion poll results.

- For example, if a confidence interval of 4 is chosen and 55% percent of the sample picks an answer we can be "sure" that if we had asked the question of the entire relevant population between 51% (55-4) and 59% (55+4) would have picked that answer.
- When we put the confidence level and the confidence interval together, we can say that we are 95% sure that the true percentage of the population is between 51% and 59% on the answer.

Think of CI as a type of variation in the confidence level (CL) that we are willing to accept for accuracy in sampled results, which is related to sample size.

The wider the CI we are willing to accept, the more variation we would expect.

- At some CI this tells us little of practical use; e.g., CI=50%
- A CI of ~0% suggests almost no variation in the results

Can we apply CIs to SCAMPs? Which CI?

Step 1: Determine OU Σ Interviewees

Use a 95% Confidence Level with a chosen CI to determine the overall Σ of sample size of interviewees across the OU.

Populations of OUs in SEI appraisals through 9/2006		95% Confidence Level							
		Confidence Interval =5%				Confidence Interval =10%			
		Low End		High End		Low End		High End	
Populations		To Be Sampled		To Be Sampled		To Be Sampled		To Be Sampled	
Low	High		%		%		%		%
14	25	14	100	24	96.00	14	100	20	80.00
26	50	24	92.31	44	88.00	21	80.77	33	66.00
51	75	45	88.24	62	82.67	34	66.67	42	56.00
76	100	64	84.21	80	80.00	43	56.58	49	49.00
101	200	80	79.21	132	66.00	49	48.51	65	32.50
201	300	132	65.67	169	56.33	65	32.34	73	24.33
301	500	169	33.73	217	43.40	73	24.25	81	16.20
501	1000	268	53.49	278	27.80	81	16.17	88	8.80
1001	2000	278	27.77	322	16.10	88	8.79	92	4.60
2001	10000	322	26.09	370	3.70	92	4.6	95	0.95
10001	or greater	370	3.7	379	1.30	95	0.95	96	0.33

What is the Population?

The target population is the entire group an appraisal team is interested in; the group (OU) about which the appraisal team wishes to draw conclusions.

But how do we determine the subsets of the OU and sample within:

- Strata
- Projects
- Interviewees

**OUs often are not homogeneous!
So we will need to perform other levels of sampling.**

Stratified Sampling

There often are factors which divide up the population into sub-populations (groups / strata) and we may expect the measurement of interest to vary among the different sub-populations. This has to be accounted for when we select a sample from the population in order that we obtain a sample that is representative of the population. This is achieved by stratified sampling.

A stratified sample is obtained by taking samples from each stratum or sub-group of a population.

When we sample a population with several strata, we generally require that the proportion of each stratum in the sample should be approximately the same as in the population.

Stratified sampling techniques are generally used when the population is heterogeneous, or dissimilar, where certain homogeneous, or similar, sub-populations can be isolated (strata); equivalent to Critical Factors.

Example

Suppose we want to understand how different project types or locations are performing CMMI across an OU. We could divide up the OU into sub-groups, take proportionate samples, and could even report by sub-group.

**We may or may not report by strata based on the appraisal contract.
But in any event, there normally are strata in OUs.**

Quota Sampling, or Non-probabilistic Sampling [DEM1]

Quota sampling is a method of sampling widely used in opinion polling and market research. Interviewers are each given a quota of subjects of specified type to attempt to recruit. For example, an interviewer might be told to go out and select 20 adult men and 20 adult women, 10 teenage girls and 10 teenage boys so that they could interview them about their television viewing.

It suffers from a number of methodological flaws, the most basic of which is that the sample is not a random sample and therefore the sampling distributions of any statistics are unknown.

How many SCAMPs have been performed using a form of Quota Sampling; e.g., pick a minimum of 3 projects?

Quota sampling needs to become history, not continued practice!

Step 2: Sampling Within Strata

1 of 3

An example spread sheet showing projects with critical factors identified:

Ordered Project Name	Size	Location	Effort Type	Project Type	Line of Business
Axxxx	S	USA	Maint	Legacy	Comm
Bxxxx	M	UK	Devel	New	Financial
Cxxxx	L	India	Serv	Legacy	Energy
Dxxxx	S	USA	Maint	Legacy	Financial
Exxxx	S	USA	Devel	New	Comm
Fxxxx	S	India	Serv	Legacy	Energy
Gxxxx	M	China	Maint	Legacy	Financial
Hxxxx	L	USA	Devel	New	Comm
Ixxxx	M	India	Serv	Legacy	Financial
Jxxxx	M	India	Serv	Legacy	Financial
Kxxxx	S	India	Serv	Legacy	Energy

1. Find the maximum variation choices within each strata.
2. Use a simple algorithm for random selections in each strata.

Step 2: Sampling Within Strata

2 of 3

3. Determine if the selections are representative of the OU %s (within strata and for OU Σ).
4. If not, continue with random selections as follows:
 - when %s are too high, discount from the last selected
 - when too low, resume selections with the chosen algorithm

Strata give insight into the minimum number of projects we should sample, especially for Focused.

Let's look at the previous table again, to discuss possible results:

1. Which strata has the maximum variation values?
2. What if we randomly chose the first three projects?
3. Have we achieved representative-ness?

Step 2: Sampling Within Strata

3 of 3

Ordered Project Name	Size	Location	Effort Type	Project Type	Line of Business
Axxxx	S	USA	Maint	Legacy	Comm
Bxxxx	M	UK	Devel	New	Financial
Cxxxx	L	India	Serv	Legacy	Energy
Dxxxx	S	USA	Maint	Legacy	Financial
Exxxx	S	USA	Devel	New	Comm
Fxxxx	S	India	Serv	Legacy	Energy
Gxxxx	M	China	Maint	Legacy	Financial
Hxxxx	L	USA	Devel	New	Comm
Ixxxx	M	India	Serv	Legacy	Financial
Jxxxx	M	India	Serv	Legacy	Financial
Kxxxx	S	India	Serv	Legacy	Energy

1. Which strata has the maximum variation values?
2. What if we randomly chose the first three projects? The last three?
3. Have we achieved representative-ness?

Step 3: How Many Σ Projects?

Using CL=95%

Total number of projects in an OU									
		CI=5%				CI=10%			
		Low End		High End		Low End		High End	
SEI Project Populations	% of Σ Number	To Be Sampled	%	To Be Sampled	%	To Be Sampled	%	To Be Sampled	%
1 - 5	9.80%	1	100.00%	5	100.00%	1	100.00%	5	100.00%
6 -10	13.40%	6	100.00%	10	100.00%	6	100.00%	9	90.00%
11-20	20.00%	11	100.00%	19	95.00%	10	90.91%	17	85.00%
21-50	2.00%	20	95.24%	44	88.00%	17	80.95%	33	66.00%
51 -100	8.70%	45	88.24%	80	80.00%	34	66.67%	49	49.00%
101-200	20.00%	80	79.21%	132	66.00%	49	48.51%	65	32.50%
201 - 500	9.50%	132	65.67%	217	43.40%	65	32.34%	81	16.20%
501 or greater	16.60%	218	43.51%	295	23.14%	81	16.17%	89	6.98%

Ok, this is not going to be accepted by most!
So what can we do?

Sampling Projects

Using CL=95%

For Focused Projects use as suggested in the Focused column below.

For Non-focused projects consider first CI=10% values.

- Or should we start with CI=5%? Cost versus increased accuracy!

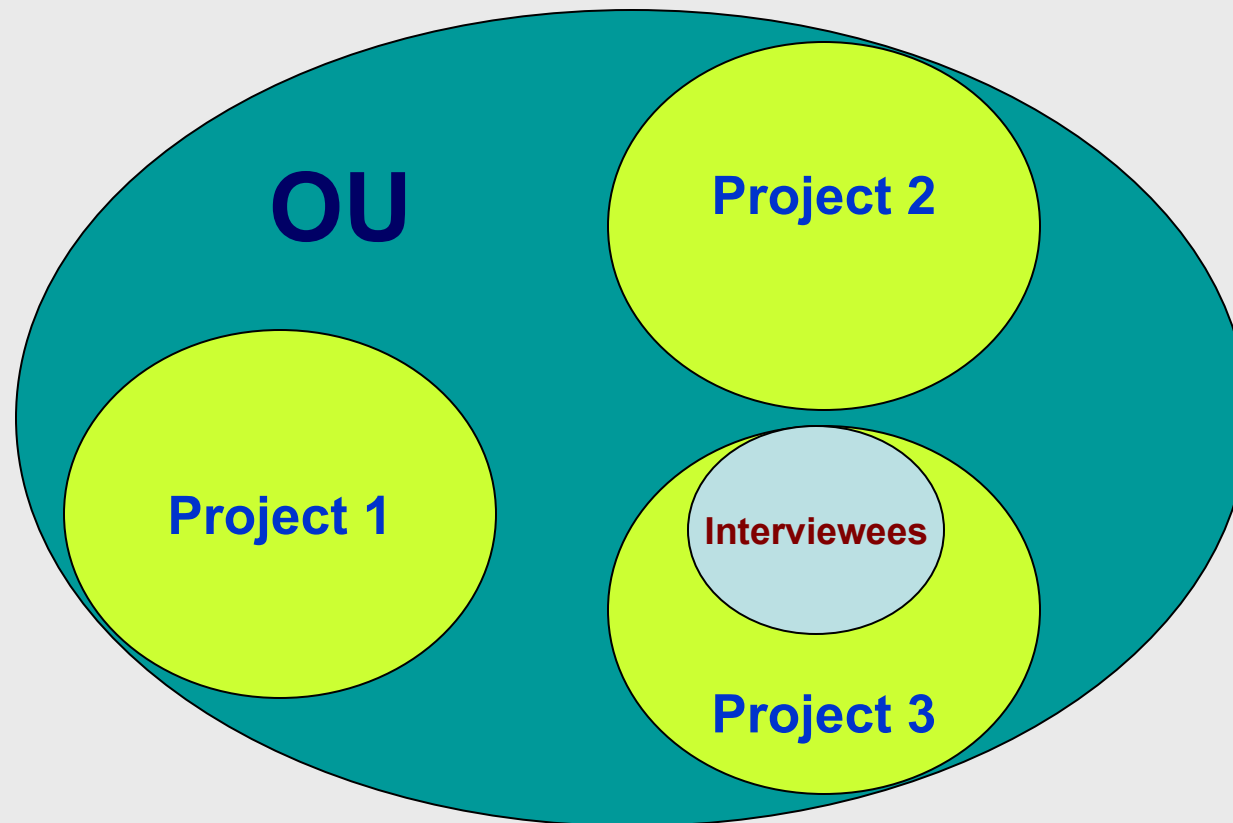
We are trying to get to a sample project set (FP+NFP) that approximates statistics for \sum of derivable sampled projects (see previous chart).

Number of projects in an OU	Focused	Non-focused		\sum of F+NF	
		CI=5%	CI=10%	CI=5%	CI=10%
1	1	0	0	1	1
2	2	0	0	2	2
3	3	0	0	3	3
4	3	1	1	4	4
5	3	2	2	5	5
6 -10	3	3-7	3-6	6-10	6-9
11-20	3-5	8-16	7-14	11-19	10-17
21-50	3-5	15-39	12-28	20-44	17-33
51 -100	5-10	40-70	29-39	45-80	34-49
101-200	min 10	70-122	39-55	80-132	49-65
201 - 500	min 10	122-207	55-71	132-217	65-81
501 or greater	min 10	208-285	71-79	218-295	81-89



Are We Done Sampling Yet?

No, we still need to pick interviewees.



Step 4: Sampling Interviewees

Simple Random Selection must be applied using a non-biased selection algorithm for FARs within sampled projects; e.g., every 3rd name.

- Recall that the projects selected demonstrate the strata are sufficiently representative

For function oriented roles within the OU; e.g., QA, SEPG, the selection may be the full set.

Thought: If there are critical factors; e.g., roles and responsibilities, these may need to be accounted for, as was done for OU strata.

Reminder:

- For SCAMPs simple random selection exists when *any* individual in the project has as likely an opportunity for being selected as any other and included in the sample.

Ordered FAR Names
Arnold Schwarzenegger
Bruce Willis
Colin Powell
Diana Krall
Elliott Ness
Fast Eddie

William Edwards Deming
Xavier Cugat
Yoland Adams
Zappa Frank

Practices and Sampling

The rules of SCAMPI characterization apply for all practices for all Process Areas within the scope of an appraisal.

How does practice performance relate to sampling results?

How do SCAMPI interviewing results relate to sampling results?

SCAMPI Rating and Sampling Integrity?

Unanswered Questions

1 of 2

SCAMPI requires rating evidence for each practice for all projects in scope:

- It does not require that all interviewees provide an answer for all relevant practices
- If, however, there is a negative response (finding) from any interviewee in any project in scope it must be captured as a weakness

We are not doing a traditional sampling analysis when doing a SCAMPI:

- 1st We say we are looking for affirmation without weaknesses in evidence, but**
 - 2nd We don't interview 100% of the OU, so we may be missing weaknesses.**
- 3rd How do we really know that the appraisal team got answers from each sampled interviewee for each practice they perform?**
- 4th An appraisal team could have only three project affirmations from which they try to derive "certainty" at the OU level on a practice.**
- 5th It is possible to only get affirmations from interviewees on only half of the practices (50% rule)**

The SCAMPI results thus could be questioned.

SCAMPI Rating and Sampling Integrity?

Unanswered Questions

2 of 2

Without asking all questions of all interviewees for all practices, we have further sub-set the sampling within the OU by practice applicability to the selected interviewees, but have we sampled sufficiently to determine a Level with certainty?

- Can we ask all interviewees about all practices that are relevant to what is performed, so we get all relevant answers?

So with SCAMPI, certainty in OU appraisal results may not be provable.

- There will always be uncertainty! But how much is acceptable?

We can approximate a more consistent set of results from appraisal teams, but this requires more guidance from SEI on sampling design!

**How much uncertainty is acceptable?
To you? To customers? To the community?**

Problems We Want To Avoid

Introducing Bias

- Picking only projects that are large
- Picking projects that are most likely to satisfy a CMMI level
- Picking too small a sample
- Projects of majority revenue only
- Selecting interviewees as representatives across projects
- Projects piloting new processes

Picking too large a sample

- Too many projects leads to higher cost and time, without a substantive increase in accuracy
- Too many interviewees
- Too many documents

Not enough onsite time to do an appraisal with integrity

Excluded projects that impact an OU's level

Inconsistent appraisals due to improper sampling

Concluding certainties from uncertain samples

Issues with Sampling

- Customers don't like / want / understand it
- LAs yield to customer pressures
- LAs may not sufficiently understand
- Cherry picking to “achieve” a Level
- Sufficient sampling rigor will not be applied to SCAMPI unless further guidance is provided by the SEI
 - Send them your cards and letters

**If we don't do it right, SCAMPI and CMMI
can become Dilbert cartoons!**

**For now start with the 4 suggested steps.
And help to find a better approach.**

THANK YOU FOR YOUR PARTICIPATION

If you are interested in working to get sampling in SCAMPI further resolved, please contact me.

When Selecting Focused Projects

- To have only one focused project is questionable practice, except when there is only 1 or 2 projects
 - SEI permits only 1
- Assume 3 Focused Projects up to a 21 project OU
- Assume 5 for 21-50
- Assume 10+ for all other OU sizes (until a better approach is determined)
- Address all critical factors
- Use simple random sampling within selected projects to pick interviewees
- Maintain focus on the four sampling principles:
 - Random-ness
 - Representative-ness
 - Timing of sample selections
 - Checking for non-bias

When Selecting Non-Focused Projects

- **5% CI (F+NF) may be more than most will tolerate, but is accomplishable, so why not use it**
- **Assume 10% CI as a minimum as worse case**
- **Address all critical factors, but now in combination with Focused Project selection**
- **Use simple random sampling within projects to pick interviewees**
- **Maintain focus on the four sampling principles:**
 - **Random-ness**
 - **Representative-ness**
 - **Timing of sample selections**
 - **Checking for non-bias**

When Selecting Interviewees

- **5% CI (F+NF) may be more than most will tolerate, but is accomplishable; refer to table in Step 1, so why not use it**
- **Assume 10% CI as a minimum as worse case**
- **Use simple random sampling within projects to pick interviewees**
- **Maintain focus on the four sampling principles:**
 - **Random-ness**
 - **Representative-ness**
 - **Timing of sample selections**
 - **Checking for non-bias**

Sample Size Calculators

There are a number of simple sample size calculators available free on the internet; e.g.,

Creative Research Systems:

<http://www.surveysystem.com/sscalc.htm>

DSS Research:

<http://www.dssresearch.com/toolkit/sscalc/size.asp>

Macoor: http://www.macorr.com/ss_calculator.htm

Wimmer Dominick Mass Media Research:

<http://rogerwimmer.com/samplesizecalculator.htm>

References

[DEM1] Some Theory of Sampling, William Edwards Deming, Dover Publications, 1966

[SEI1] Standard CMMI Appraisal Method for Process Improvement (SCAMPI) A, Version 1.2; Method Definition Document

[THOM1] Sampling, Second Edition, Steven K. Thompson, Wiley Series in Probability and Statistics, 2002

[WIK1] Wikipedia