Sampling for More Effective & Efficient Program/Project Progress Reporting [1]

Dr. Kenneth F. Smith, PMP

Last month, I reviewed the basic precepts for **random sampling**, as well as possible pitfalls to avoid or mitigate when collecting data. However, with those criteria aforethought, undertaken with prudent precautions, sampling is an economical, efficient and effective means for acquiring useful information about a target population, as well as a program/project's interventions to improve a situation. Consequently, throughout my international economic, infrastructure and social development career in program and project management in diverse sectors, whenever I encountered dysfunctional census-type management information systems (MIS); I sought to supplement – if not supplant -- them with some systematic sampling capability.

This month I continue on that tack by highlighting the weaknesses inherent in census-style recurrent reporting, and -- even if not replacing the extant MIS in its entirety -- offering **an even '*quicker & easier*' sampling approach** to validate program / project performance data.

Unquestionably, a lot of data are needed to measure implementation performance and support program / project management decision-making. Unfortunately, all too often the structures institutionalized to acquire and analyze data are bureaucratically burdensome, cumbersome, costly, excessive, untimely and – *regrettably* -- ultimately ineffective.

At the request of one government national agency Executive – *for example* -- I interviewed, reviewed, categorized and compiled the following '***fourteen faults of recurrent reports***' within his agency for his follow-up administrative-cleanup action.

Fourteen faults of recurrent reports

1. Many different reports, formats and structures exist for managing the same program.
2. Extensive and excessively detailed data are required.
3. There are numerous redundancies in data requirements, some obvious gaps in coverage, ambiguity in requests for data and responses; and (in some instances) discrepancies exist in the answers provided.
4. The authenticity (validity and accuracy) of the data in many instances is unknown.

[1] How to cite this article: Smith, K. F. (2022). Sampling for More Effective & Efficient Program/Project Progress Reporting, *PM World Journal*, Vol. XI, Issue II, February.

5. National targets are set by a "Top Down" Process, and assigned to Provinces without their participation, request for concurrence as to realism, or feasibility to accomplish.
6. Reporting requirements are too many, and too frequent for the field staff to fulfill meaningfully.
7. Most reporting requirements consume excessive amounts of the field technician's time to complete.
8. Most reports from the field are submitted too late to be of use for action-oriented decision-making at intermediate and higher levels.
9. Intermediate levels as well as the Central Office spend inordinate amounts of time and effort summarizing data and preparing reports for their level of management.
10. Some data are summarized manually (primarily in spread-sheet form) and laboriously retyped or re-entered, with many errors. [Late or missing data is simply ignored. Missing or delinquent reports are seldom followed up.]
11. Many data reported are *ad hoc* and *ad hominem.* Reports are largely descriptive narrative with some statistics scattered 'randomly' throughout the text.
12. Data are seldom analyzed by the staff or management at any level. For the most part, there is no standardized data base or on-going measurement against targets for performance accomplishment, or trend analysis.
13. Almost every request by managers for information generates a frantic new search for -- or re-creation of -- historical data, and often results in the creation of new reporting requirements.
14. Lack of feedback. Data reporting is a "one way street." Data eventually comes in to the Central Office, but only directives go out. No indication is made of the utility of the data, nor are summaries of information provided to the field for their possible use. Neither are reported problems dealt with in any systematic manner, or feedback provided.

While, today, computerization can readily resolve some of these syndromes, nevertheless, the litany of many of the less-than-desirable issues still linger on in many programs and projects; especially reporting bias by individuals with vested interests in the outcome. However, surveys can bypass – if not obviate -- bureaucratic bungling from endemically-entrenched and regularly-reported recurrent census-style service statistics; as well as clarify obfuscated issues, &/or worse – ***reveal deliberate distortion*** by program/project operational personnel.

In monitoring and evaluating program and project performance, the three key aspects to assess are:

1. **Service Delivery** — Have the services planned actually been **provided?** *[i.e. technical support: physicians, nurses, midwives, agricultural extension agents; information / education campaigns, training; infrastructure, clinic facilities, roads, equipment, medications, fertilizer, etc.,]*
2. **Technical "Quality"** — Are the services being used/implemented **correctly**?
3. **Target Beneficiary "Coverage"** — Are the target beneficiaries receiving, using, &/or showing evidence of **benefitting** from the program or project's services?

Random Sampling is generally preferable to a **Census**[2] for monitoring & evaluation because sampling is flexible, less biased, provides *reasonably accurate* "order of magnitude" data on key indicators in a shorter time, and for considerably less effort and cost than a structured periodic recurrent reporting system. **Periodic random sample surveys** conducted by independent teams can gather data on key indicators far more efficiently, effectively, **and moreover –** *given probability considerations incorporated in sampling frames* **-- more accurately** for centralized management analysis and use.

Nevertheless, as indicated last month, random sampling still requires a large number of samples — *sometimes several hundred* — depending on the nature of the indicators; the extent of variability of the population; and the manager's need for accuracy. Thus, where there is relative inaccessibility to the target population due to the program or project's geographic coverage, work environment, scarce resources &/or inadequate transportation for data collectors, random sampling can still pose a formidable challenge for data collection and analysis.

However, if the Manager can pre-establish ***Technical Quality Performance Standards*** for each of the indicators in terms of **means** or **percentages** *and* merely wants to know ***whether or not those standards have been, or are being achieved*** -- rather than all the dirty details[3] -- **random sampling is still not the best approach.** In such situations he/she can borrow a **NON-RANDOM SAMPLNG technique** from our **Quality Control (QC)** colleagues; as their ***'BECAUSE'*** approach **-- *a*ctually 'BQAS' Batch Quality Acceptance Sampling --** ***can satisfy the need much more efficiently and effectively with considerably fewer samples.***

Essentially, for quality control purposes, **additional items of a particular product *should be uniform* –** or within tolerable range acceptance standards. ***Exceptions are anomalous!*** So, **any items** from a **product population** -- or smaller **batch** from a production run of that population – ***should be identical, or with minimal acceptable variance from that standard.*** Should that error rate be exceeded, the entire batch from which the sample was drawn, is rejected.

[2] i.e. *100% recurrent pre-formatted reporting* from operational levels to the Project Manager

[3] Details such as national program "estimated coverage is 75 percent, ranging from 30 to 90 percent by area;" or "the average farmer's rice harvest is 45 cavans/hectare, ranging from 10 to 140 cavans/hectare;" or whatever.

In essence, BQAS dichotomizes performance as either *"Satisfactory"* or *"Unsatisfactory"* based on professional judgement whether predetermined *Standards* are being met in a particular case; but *the precise range or performance level is not captured or recorded*.

Thus, with the **Batch Quality Acceptance Sampling (BQAS)** approach, ***large samples*** **-- i.e. 1,500 or even 100 --** ***are not necessary to determine whether or not a QUALITY STANDARD has been achieved for a much larger program or project***. Indeed. **as few as 19 non-random samples can give results with relatively high accuracy**; **while 28 non-random samples[4] reduces *risk levels* to within 4 percent.**

Similarly, small non-random sample results obtained from various geographic areas (such as barangays or municipalities) can be aggregated to get a broader picture of performance on a provincial, regional or national program or project with respect to specific indicators.

To apply the BQAS, a "Triage" approach must be designed, as follows:

1. **DEFINE THE INDICATOR STANDARD — in percentage terms — for Three Performance Levels[5]:**
    1. **Satisfactory** — *[*i.e. Health Example: *80 percent or more of children under one year of age are breastfed ;* Agriculture Example: *80 percent of farmers are using high yielding variety (HYV) seed]*
    2. **Poor** — *[i.e.* Health: *51 to 79 percent . . . breastfed ;* Agriculture: *. . . HYV]*
    3. **Unsatisfactory** — *[i.e. 50 percent or less . . . breastfed; . . . HYV]*
2. **DETERMINE THE TOLERABLE ERROR in Classifying Performance as "Satisfactory" or "Unsatisfactory" —** *[i.e. +/-5% = 2 standard deviations]*
3. **DECISION-RULE CRITERIA Determine:**
    - **Appropriate Sample Size** A judgement call — *for example: 28* **—** based on feasibility to collect data as well as the need for accuracy (as determined by the Standards for the three levels, and the Project Manager's Tolerable Error)

[4] Allowing for 9 "Unsatisfactory" cases**.**

[5] NOTE: The actual percentages appropriate for each Standard should be determined by one or more stakeholders — i.e. the project manager, the technical professional staff, a technical consultant "judge", or a "jury" composed of stakeholders, professionals, and target beneficiaries

- **How many instances** ***which do not meet the standard*** **will be tolerated in the sample** ***and the batch still accepted*** — *for example: 9* — based on need for accuracy (as determined by the Project Manager's Tolerable Error)

All three of these step-decisions are interrelated in the formula below:

Probability of Encountering an Unsatisfactory Instance in a Sample	$= \frac{n!}{[a!\,(n-a)!]} \times p^a \times q^{n-a}$

Where:

n = Sample Size

! = "Factorial" i.e. 7 x 6 x 5 x 4 x 3 x 2 x 1

a = # Satisfactory Instances in the Sample

n - a = # Unsatisfactory Instances in the Sample

p = Satisfactory Standard (Expressed as a decimal % *i.e. .8 = 80%*)

q = % Below Standard *i.e. (1 - p) Also e*xpressed as a decimal %

Solving this and related equations is extremely time consuming and subject to error. **I therefore developed a couple of illustrative templates for reference and interactive Templates to compute the answers.**[6]

[6] NOTE: The underlying computations for BQAS are much too burdensome for most people to plough through, so you are not expected to slavishly study and acquire skill in applying them. They are for reference in my PRAXIS book with examples for anyone seeking a deeper understanding of the basis for the BQAS process and resultant data in the tables. Suffice it to say, my templates have the formulas embedded, and do all the "*number crunching*" for you.

Figure 1

The "TRIAGE" System *2020 1997 Dr. Kenneth F. Smith, PMP*

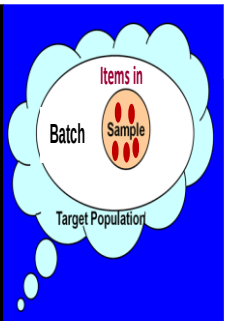
QUALITY PERFORMANCE STANDARDS, & DECISION-RULE CRITERIA for SAMPLING, and RISK ASSESSMENT

EXAMPLES of RISK, GIVEN VARIOUS SAMPLE SIZES and DECISION-RULES

SAMPLE SIZE #	SAMPLE # *BELOW* STANDARD #	STANDARD "SATIS" (AT LEAST) %	STANDARD "UNSATIS" (LESS THAN) %	PROVIDER "SATIS" ACCURACY %	"SATIS" RISK %	CONSUMER "UNSATIS" ACCURACY %	"UNSATIS" RISK %	TOTAL RISK %
3	1	95%	30%	99%	1%	78%	22%	22%
6	1	90%	60%	89%	11%	77%	23%	35%
6	1	85%	50%	78%	22%	89%	11%	33%
6	1	80%	40%	66%	34%	96%	4%	39%
7	2	80%	40%	85%	15%	90%	10%	24%
8	2	80%	50%	80%	20%	86%	14%	35%
9	3	80%	50%	91%	9%	75%	25%	34%
10	3	80%	50%	88%	12%	83%	17%	29%
11	3	80%	50%	84%	16%	89%	11%	27%
12	4	80%	50%	93%	7%	81%	19%	27%
13	4	80%	50%	90%	10%	87%	13%	23%
14	4	80%	50%	87%	13%	91%	9%	22%
15	5	80%	50%	94%	6%	85%	15%	21%
16	5	80%	50%	92%	8%	89%	11%	19%
17	5	80%	50%	89%	11%	93%	7%	18%
18	6	80%	50%	95%	5%	88%	12%	17%
19	6	80%	50%	93%	7%	92%	8%	15%
20	6	80%	50%	91%	9%	94%	6%	14%
21	7	80%	50%	96%	4%	91%	9%	14%
22	7	80%	40%	94%	6%	99%	1%	6%
23	7	80%	50%	93%	7%	95%	5%	12%
24	8	80%	50%	96%	4%	92%	8%	11%
25	8	80%	50%	95%	5%	95%	5%	10%
26	9	80%	50%	98%	2%	92%	8%	11%
27	9	80%	50%	97%	3%	94%	6%	9%
28	9	80%	50%	96%	4%	96%	4%	8%
29	9	80%	50%	95%	5%	97%	3%	8%
30	10	80%	50%	97%	3%	95%	5%	7%
50	16	80%	50%	99%	1%	99%	1%	2%
100	31	80%	50%	100%	0%	100%	0%	0%
150	43	80%	50%	100%	0%	100%	0%	0%
200	55	80%	50%	100%	0%	100%	0%	0%
250	67	80%	50%	100%	0%	100%	0%	0%
300	78	80%	50%	100%	0%	100%	0%	0%
500	124	80%	50%	100%	0%	100%	0%	0%

Figure 2

The "TRIAGE" System for Batch Quality Acceptance Sampling



QUALITY PERFORMANCE STANDARD SETTING, RISK LEVEL ASSESSMENT & DECISION-RULE CRITERIA for NON-RANDOM 'SPOT-CHECK' SAMPLING

TRIAL & ERROR: Adjust the Standards, Accuracy & Sampling Criteria below to whatever is Pertinent in your Situation
Enter your values in the YELLOW cells; **then APPLY DECISION RULE 1 or 2**

BATCH SIZE	50	*[Actual, or best Estimate]*		**Standard & Risk Level Decision Rule Test**		**TARGET POPULATION COHORT in the BATCH**
IF QUALITY PERFORMANCE STANDARDS are			**Accuracy is**	***Recommend at least 95 % Level***	**Desired %**	
1. At Least	80	% *"SATISFACTORY"* @	91%	*Too Risky : TRY AGAIN*	95	40 Satisfactory
& 2. Less than	5	% *"UNSATISFACTORY"* @	100%	*Acceptable Risk Level*	95	2 Unsatisfactory
With the RESIDUALS being of *'POORER'* Quality than "*SATISFACTORY*," but not completely *"UNSATISFACTORY."*						9 Between S & U
DECISION RULE: 1. ACCEPT the ENTIRE BATCH as *MEETING the above STANDARDS*						**IF BATCH IS ACCEPTED**
Given the RISKS **WHEN**	6	**(or less) Spot-checked items in the Sample *within the BATCH***				***Subject to GREY AREA RISK***
Below **in a SAMPLE of**	20	***are Deemed to be***	**"UNSATISFACTORY"**			**Range Possibilities below:**
Type 1 Error: False NEGATIVE	***PROVIDER Risk: UP to***	9%	**of the Batch is estimated to be "UNSATISFACTORY," but could Actually be "SATISFACTORY."** ***So, the "SATISFACTORY" Level could be At Least***		89%	44 >Satisfactory
GREY AREA RISK =	9%	***"In-between" items/cases/situations in the BATCH may be INACCURATELY rated as either "UNSATISFACTORY" or "SATISFACTORY"***				***as the "BEST CASE;" or the "WORST CASE" being:***
Type 2 Error: False POSITIVE	***CONSUMER Risk: UP to***	0%	**of the Batch is estimated to be "SATISFACTORY," but could Actually be "UNSATISFACTORY."** ***So, the "UNSATISFACTORY" Level could be As Much As***		4.9%	1 >Unsatisfactory
2. BUT *REJECT the ENTIRE BATCH*		**IF**	7	***or More are "UNSATISFACTORY"***		3. Grey Area Ranges are adjusted if ***Provider*** or ***Consumer Risk Exceeds 100%***

*NOTES: 1. "Grey Area" Risk may not always equal "Provider + Consumer" Risk because Percentages are rounde 2. Also, due to rounding, 0 & 100% may also be imprecise.

Figure 3

PROJECT QUALITY PERFORMANCE -- AREA NUMERICAL ASSESSMENT

Non-Random Batch Quality Acceptance Sampling (BQAS) TRIAGE SYSTEM

© Dr. Kenneth F. Smith, PMP November 1997 Lotus123; Excel September 2005

INDICATOR Description: i.e. Percentage of Children Under 3 years of Age who have completed a series of DPT Vaccinations

Sample - Hypothetical Data for Illustrative Purposes Only

1	2	3	4	5		6		7		8		9	10	11	12
	POPULATION BASE		STANDARD Criteria			DECISION RULE				FINDINGS		PROJECT PERFORMANCE SUMMARY EVALUATION			Weighted
PROJECT UNIT TARGET AREA, or PROJECT ACTIVITY being ASSESSED	TARGET POPULATION SIZE	RELATIVE Wt of UNIT to TOTAL TGT POP	INDICATOR SATISFACTORY LEVEL QUALITY PERFORMANCE STANDARD	PROVIDER ACCURACY computed from the cumulative binomial formula	Accy check	SAMPLE SIZE	Error check	ACCEPTABLE NUMBER BELOW STANDARD	Error check	ACTUAL NUMBER BELOW STANDARD	Error check	STATUS & % UP TO STANDARD	NUMBER of PROJECT UNITS Up To STANDARD	MINIMUM PERFORMANCE LEVEL OF UNIT	% POPULATION Satisfactorily SERVED
		a / Sum a										SAT if g <= f		c+ %dev g% fr f%	b, if h = SAT
	a	b	c	d		e		f		g		h	I	i	k
			PROF Judgement	At least 95 % for 2 SDs		What is Feasible?		Minimum # for Accy							
Barangay 1	600	21%	80%	98%	OK	18	OK	7	OK	8	OK	UNSAT	0	71%	0%
Barangay 2	170	6%	80%	95%	OK	18	OK	6	OK	8	OK	UNSAT	0	63%	0%
Barangay 3	230	8%	80%	97%	OK	20	OK	7	OK	7	OK	SAT	1	80%	8%
Barangay 4	84	3%	80%	98%	OK	15	OK	6	OK	6	OK	SAT	1	80%	3%
Barangay 5	257	9%	80%	99%	OK	8	OK	4	OK	2	OK	SAT	1	100%	9%
Barangay 6	149	5%	80%	98%	OK	6	OK	3	OK	4	OK	UNSAT	0	47%	0%
Barangay 7	203	7%	80%	97%	OK	7	OK	3	OK	1	OK	SAT	1	100%	7%
Barangay 8	190	7%	80%	98%	OK	15	OK	6	OK	7	OK	UNSAT	0	69%	0%
Barangay 9	130	4%	80%	99%	OK	8	OK	4	OK	5	OK	UNSAT	0	55%	0%
Barangay 10	142	5%	80%	96%	OK	14	OK	5	OK	7	OK	UNSAT	0	58%	0%
Barangay 11	178	6%	80%	95%	OK	25	OK	8	OK	3	OK	SAT	1	100%	6%
Barangay 12	161	6%	80%	97%	OK	13	OK	5	OK	4	OK	SAT	1	93%	6%
Barangay 13	169	6%	80%	96%	OK	17	OK	6	OK	4	OK	SAT	1	98%	6%
Barangay 14	245	8%	80%	95%	OK	33	OK	10	OK	3	OK	SAT	1	100%	8%
TOTAL	2,908					217						57%	8	80%	53%

NOTE: Percentages are rounded, so may not total exactly. 0% = 0 -->0.049 % & 100% = 99.5 % --> 99.99 %

Standards should be ***realistic***[7] — **even set *low* initially** (so as not to overwhelm implementers at first, and to allow them room for growth) — for ***the* Satisfactory** *and* **Unsatisfactory categories;** then successively raised as major problems are detected, corrective action taken, and improvements expected. Differing *management-determined* values are permitted for the extreme ranges. However, each case selected in the sample is judged as ***only one of two* states** — ***either "Satisfactory" or "Unsatisfactory".*** Thus, cases falling in the intermediate "grey area" — will be ***over-rated or under-rated*** — as either *a* ***"Consumer"*** *or a* ***"Provider"*** risk.

- **CONSUMER RISK** Is the risk that the situation is ***Actually Below Standard*, but is rated as "Satisfactory".** [False Positive.]

- **PROVIDER RISK** Is the risk that the situation is ***Actually Above Standard*, but is rated as "Unsatisfactory".** [False Negative.]

  Provider Risk causes three major problems for Project Managers:

  1. **Resource Overkill** Unnecessary allocation or reallocation of resources — to correct a perceived problem situation which does not really exist. *[Conversely, with Consumer Risk, resources are not provided (or withheld and reallocated to other uses) because they are presumed to be no longer required.]*
  2. **Internal Demoralization** Project Staff become resentful when the Project Manager incorrectly perceives them as inadequate performers — and they may even reduce the quantity and quality of their work because they feel unappreciated
  3. **External Criticism** Stakeholders — and others — will unduly criticize the Project's Managers for perceived poor performance.

Both types of risks are valid concerns for the project manager *and should be expected*. However, *low levels of misclassification are* ***not*** *a problem;* as eventually both will be rectified if a conservative *"better safe than sorry"* management approach is adopted by sampling to monitor and reduce ***Provider*** Risk. [Direct Target Beneficiary feedback should pinpoint specific Consumer Risk instances that were misclassified by sampling.]

OVERALL PROJECT PERFORMANCE can then be assessed relatively easily based on aggregated analysis of sub-area results.

Hopefully, with awareness from this article, some readers will be encouraged to add BQAS to their 'DIY' sampling 'toolkit' &/or seek help from their Quality Control colleagues to improve their program and project monitoring & performance evaluation practices.

[7] Realistic, given the working environment and resources. Not *"Benchmarks"* — i.e. attempting to emulate the "best in the industry."

About the Author

Dr. Kenneth Smith

Honolulu, Hawaii
& Manila, The Philippines

Initially a US Civil Service Management Intern, then a management analyst & systems specialist with the US Defense Department, Ken subsequently had a career as a senior foreign service officer -- management & evaluation specialist, project manager, and in-house facilitator/trainer -- with the US Agency for International Development (USAID). Ken assisted host country governments in many countries to plan, monitor and evaluate projects in various technical sectors; working 'hands-on' with their officers as well as other USAID personnel, contractors and NGOs. Intermittently, he was also a team leader &/or team member to conduct project, program & and country-level portfolio analyses and evaluations.

Concurrently, Ken had an active dual career as Air Force ready-reservist in Asia (Japan, Korea, Vietnam, Thailand, Indonesia, Philippines) as well as the Washington D.C. area; was Chairman of a Congressional Services Academy Advisory Board (SAAB); and had additional duties as an Air Force Academy Liaison Officer. He retired as a 'bird' colonel. After retirement from USAID, Ken was a project management consultant for ADB, the World Bank, UNDP and USAID.

He earned his DPA (Doctor of Public Administration) from the George Mason University (GMU) in Virginia, his MS from Massachusetts Institute of Technology (MIT Systems Analysis Fellow, Center for Advanced Engineering Study), and BA & MA degrees in Government & International Relations from the University of Connecticut (UCONN). A long-time member of the Project Management Institute (PMI) and IPMA-USA, Ken is a Certified Project Management Professional (PMP®) and a member of the PMI®-Honolulu and Philippines Chapters.

Ken's book -- **Project Management PRAXIS** (available from Amazon) -- includes many innovative project management tools & techniques; and describes a "**Toolkit**" of related templates available directly from him at kenfsmith@aol.com on proof of purchase of PRAXIS.

To view other works by Ken Smith, visit his author showcase in the PM World Library at https://pmworldlibrary.net/authors/dr-kenneth-smith/