Efficiently Limiting Census Errors When Quality Control Parameters Range Freely

Glenn Wolfgang
U.S. Census Bureau

This report is released to inform interested parties of ongoing research and to encourage discussion. The views expressed on statistical, methodological, or operational issues are those of the author and not necessarily those of the U.S. Census Bureau.

Keywords: quality management, acceptance sampling, inspection efficiency

1. Introduction

Assuring quality, managing requirements, and developing process metrics are common themes in project management and software engineering. This paper focuses on practical quality control (QC) strategies to manage and measure quality relative to a quantified requirement in census data keying and response coding. These strategies build on statistical QC techniques developed over decades. They tie acceptance sampling plan parameters of the QC operation to specific output quality requirements. They manage situations where it is not feasible to prescribe or control lot size or sample size to constant values. They guide the selection of units for QC inspection to assure that those quality goals are met while minimizing inspection workloads and the operational effort needed to implement the procedures.

This paper discusses three criteria for evaluating QC procedures:

- Effectiveness -- Does the QC procedure assure meeting the error limit requirement?
- Efficiency -- Does the QC procedure minimize inspection workloads?
- Burden -- Does implementing the QC procedure complicate or increase the effort of handling data?

The general strategies for optimizing QC procedures with respect to these criteria are being developed in the context of operations where humans capture respondents' responses in a useful electronic form. Data keying is keyboard entry of answers to check-box or write-in responses on returned paper questionnaires. Some keying operations are highly automated using scanned images. Others are based on keying directly from paper forms. Response coding is the assignment of one or a set of three-character alphanumeric codes for an electronically captured write-in response to a data element (such as race or ancestry). The codes are more useful for summary or analysis than the original idiosyncratic character write-in strings.

Managing quality in these operations involves classic statistical quality control, specifically, acceptance sampling methods. The routine focus is on the quality of products, rather than the process of the operations, although diagnosing and improving the process is a valued follow-up goal motivating the inspection of products. Traditional statistical QC methods are particularly suited to repetitive operations where the result consists of multitudes of individual product units, an error or deficiency could occur in any unit, and the correctness of each unit is important.

For data keying, a key field, a single answer space or checkbox containing a response to be captured off a paper form, is the product unit. In response coding, the unit is a response code or set of codes assigned to a unique write-in response. In each operation, these units are organized naturally in batches, which are formed for convenience and minimizing risk in handling and tracking work through many processing steps. In keying, a batch is a stack of forms sent together through processing. In response coding, a batch is a collection of write-in fields transmitted to expert coders in a single computer file. The collection of all product units in a batch comprises a lot in the traditional QC terminology. The terms lot and batch are thus used interchangeably in this context.

2. Objectives of the Data Keying and Coding QC

A general objective of census data keying and response coding QC operations implemented in intra-decade census tests is to assure that the amount of error in keying or coding results does not exceed one percent. That rate is specified in written requirements established before the operations. Another important general objective is to provide timely, effective feedback for continuous improvement of the keying or coding process. To carry out these general objectives, the QC procedures are designed specifically to:

- Monitor errors during ongoing production,

- Replace identified errors with corrections,
- Inspect batches completely when there are too many errors to satisfy the one percent error limit requirement,
- Identify concentrations of errors, often due to an individual keyer's performance, which could be improved by intervention (such as retraining or reassignment) to eliminate the error source,
- Alert supervisors to general process problems, which could be improved by intervention (such as procedure revision),
- Minimize QC inspection workloads, and
- Minimize overall operational burden.

All these objectives shape the operational features of a QC program. The focus of this paper is limited to the statistical quality control criteria, effectiveness and efficiency, and on burden to the extent that burden impacts those criteria. Continuous improvement objectives, where supervisors diagnose and intervene on the production floor by adapting keyer or coder procedures or retraining for process improvement purposes are important but not addressed in this report.

3. Statistical QC Methods

3.1. Verification and Adjudication

Keying and coding operations involve independent verification of production results with adjudication. First, a production keyer or coder produces data for every unit in a batch. A different person, in the verification role, independently keys or codes designated units in that batch. Usually, only a sample of batch units is verified. However, as described below, sometimes there is a complete, full verification of that batch, so every unit in the batch is verified in the end.

When production and verification values do not agree, a skilled, experienced worker, in the adjudication role, reviews the differing results, records the correct result, and records whether the production value was in error. If the number of production errors in a verification sample is higher than a specified criterion, the batch is rejected and recycled for full verification and adjudication. Whenever inspection leads to correcting a production result, the correction is transmitted in final data. This is called rectification.

3.2. Selecting Sampling Plans

Outcomes of verification and adjudication, such as the rate of outgoing error and the amount of additional inspection resulting from batch failures, depend on a set of three parameters, known as a sampling plan -- lot or batch size (N), sample size (n), and an acceptance number (c). A sampling plan is applied each time a batch is verified. The count of fields that need to be coded or keyed in those batches is N, of which n are initially sampled for QC inspection or verification. No more than c errors are acceptable; otherwise the batch fails. The term sampling plan may denote everything one does to conduct acceptance sampling QC for an operation. The parameters, which are the essence of this general definition, are all the term implies in this paper.

In manufacturing settings, it is common to choose a single sampling plan to be applied to all production. Using statistical tools discussed below or with analysis of the performance of the sampling plan, the outcomes can be predicted and fine-tuned to suit production goals.

It is not feasible to force keying or coding operations into a one-plan approach. Batch sizes are dependent on operational considerations like the number of forms that come in for processing or how many write-in responses happen to need coding on that day. There are many unpredictable variables like how many items are answered on each questionnaire. It would take extraordinary efforts, like pre-counting key fields or restructuring batches to force them all into lots of exactly N units. It also would not likely work. The consequent extra work, confusion, and other problems would compound into lost data or other errors. That unnecessary, risky, and costly extra effort illustrates the concept of burden.

The only operationally feasible way to handle keying physical forms is to key the whole questionnaire. Asking keyers to pick and choose fields on a form for sample keying distracts and confuses natural keying rhythms, introducing error and other burden. Keying complete forms limits the distraction to moments between finishing one form and beginning another. So without splitting physical forms between lots or sample status, sample size is as hard to control as batch size. Somehow, sampling plans have to adapt to the values of N and n that arise in naturally occurring batches.

A common approach to selecting a sampling plan involves simply sampling every tenth field that comes into production. The values of N, n, and c may, by chance, comprise an effective or efficient plan, but there is no guarantee of that.

Therefore, the task is to identify, for each batch encountered, the sampling plan with optimal effectiveness, efficiency, and least burden. The first step is to identify plans that are effective in limiting error at the required level. Section 3.3 describes calculations used to do that. For a given lot size or sample size, there are many sampling plans that can be shown by these calculations to be effective in limiting error as required.

The next step is to select, from the variety of effective plans, an operationally efficient sampling plan, which minimizes inspection workload, that is, requires the fewest production units be inspected. Section 3.4 describes calculations used to do that.

Throughout those steps, there is a general aim to avoid operational complications or management burden. Software systems may be designed to facilitate the cognitive task, avoid distractions, and minimize human data handling. While this report does not explore those considerations, it is important to evaluate different mechanisms for picking an effective and efficient plan that will not increase delays or other burden. Section 3.5 explores varieties of sampling plan strategies.

3.3. Identifying Effective Sampling Plans

An acceptance sampling QC methodology based on the Average Outgoing Quality Limit (*AOQL*) statistic is well suited to these operations, since they are based on rectification and focused on outgoing quality. This method assures that, in the long run, the error rate is not worse than the specified limiting *AOQL* value, no matter what level of error exists in product submitted for inspection (Dodge, 1963). To be effective in assuring that the rate of final product error is below a specified limit, the sampling plan for each of many production batches needs to have an *AOQL* below that rate.

In the *AOQL* approach, a stream of product units is organized into lots (of size N). Generally a lot consists of all the units in a batch, as defined by the operation. A sample (of n units) is inspected to determine if the number of defective units or errors is within acceptable limits (equal to or less than c). If the batch is acceptable, it goes to the customer. All units in rejected batches are inspected and corrected before going to the customer. In an *AOQL*-based procedure, only unsampled units of accepted batches could contain uncorrected errors going out to the customer.

Computations for the *AOQL* are built upon the sampling plan parameters, N, n, and c, as well as p, the proportion of error in the production units before inspection and correction. This error rate is called process error or incoming error. While p is unknown, any prior information convincingly suggestive of its approximate value may be useful for optimizing efficiency. Yet for effectiveness in meeting an error limit requirement, the level of p cannot be assumed.

An understanding of the *AOQL* is the basis for options in efficient, effective control of errors. Wadsworth et al. (2002, pp 520-526) and Montgomery (2001) describe the *AOQL* more completely.

At the core of the *AOQL* are hypergeometric probabilities (Hg), which give the chances that a lot of N units containing D defective units and a sample of n units, will yield d defective units in the sample. The unknown p is D/N, so $D = pN$, approximately. If d is one of the values acceptable given the sampling plan's limit c, the probability of accepting a lot with specific sampling plan and incoming error values N, n, c, and p, may be expressed as the probability of acceptance (Pa):

$$Pa = \text{sum over } d = 0 \text{ to } c \ (\text{Hg}\ (N, pN, n, \text{and } d)). \qquad (1)$$

The *AOQ* is the projected rate of outgoing error based on the sampling plan and a presumed p given two operational rules:

- Failed batches are fully (one hundred percent) inspected, and
- All units inspected, whether in the sample or in batches rejected, are rectified.

That means all errors identified are corrected. Since there are then no errors left in previously failed batches or samples, outgoing error is found, at a rate of p, in only the non-sampled portion, $(N-n)/N$, of the Pa batches that were accepted. So,

$$AOQ = p * ((N-n)/N) * Pa \qquad (2)$$

Since we never know the true p, we don't know which *AOQ* applies to a given batch. At $p = 0$ there is no incoming error to pass on as outgoing error. At very high values of p, all batches are likely to be rejected and fully rectified and thus pass on no outgoing error. Between those values, the *AOQ* rises to a maximum and then falls off.

$$AOQL = \text{max over } p\ (AOQ) \qquad (3)$$

The *AOQL* approximates how bad it could possibly get, not how bad it will likely get. Also, it is possible, but quite unlikely, that a particular batch contains more outgoing error than its *AOQL*. That would most likely happen if the incoming error is near the level of p for which the *AOQL* is defined, and the sample happens to under-represent the batch error. If that were to occur in any of the batches, the other batches compensate for it by their relatively small contributions to overall outgoing error. So the *AOQL* seems to be a safe measurement tool to assure a specified error limit. All that is needed is to apply a sampling plan that has an *AOQL* fitting the requirement. If the requirement is for "at least 99 percent accuracy" or "no more than one percent error," any sampling plan with an *AOQL* of 0.01 or lower is effective.

3.4. Selecting Efficient Sampling Plans

Average Total Inspection (*ATI*) is a traditional metric to help evaluate workloads of sampling plans (Wadsworth et al., 2002, p 527). It projects the average number of cases inspected beyond production in a batch with a given sampling plan. The *ATI* is the sample size plus the portion *(1 - Pa)* of non-sample units *(N - n)* units that are inspected, assuming the sample units are not inspected again after a batch is rejected.

$$ATI = n + (1 - Pa)(N - n) = N * (1 - AOQ / p) \quad (4)$$

The value of *ATI* is specific not only to a given sampling plan (with its *N*, *n*, and *c* values) but to the value of *p* on which *Pa* is computed. Transforming *ATI* to a percent permits comparing workloads among batches with different parameters. Expressing it as a function of *p* and *AOQ* lets it show the impact of incoming error on workload. The Inspection Percent (*IP*) captures these two goals: measurement of plans' expected efficiency that is both standardized and differentiable relative to *p*.

$$IP(p) = 100 * ATI / N = 100 * (1 - AOQ / p) \quad (5)$$

While the incoming error *p* is unknown for each individual batch, past or current experience can provide very good clues to the approximate level of incoming error. Any relevant prior information about the approximate value of *p* might be useful in selecting sampling plans that would optimize workload efficiency. If selection of an error limit requirement implies some prior information or interest in a level of *p*, that suggests a default *p* value for *IP*(*p*), equal to the specified limit. If past experience suggests strongly that error rates are going to be much less than that limit, the *IP*(*p*) for that *p* is more relevant to efficiency evaluations.

3.5. Sampling Plan Strategies

For each QC operation, planners may develop a strategy to identify an effective and efficient sampling plan. Something that can be implemented without delay or distraction during production is most valuable. A strategy is a mechanism to identify a sampling plan using what is known about requirements, values of *N* or *n*, and other conditions of the operation. Strategies fall into one of two patterns. Lot-size strategies name one sampling plan for each value of *N* in range, as described in sub-section 3.5.1. Sample-size strategies identify sampling plans primarily by *n*, as discussed in sub-section 3.5.2.

Strategies vary endlessly, partly because of the variety in operational characteristics and requirements to which they are fitted. One operation might call for an error limit requirement at two percent while another requires a half percent. An operation may prohibit c = 0 because production staff respond better when they know they will not "fail" for making just one error. Batch sizes might naturally range from 1000 to 2100 for one production task, while 50 to 100 fits another better. There might be good reason to presume an exceptionally low incoming error rate, even while QC is still needed to check for batches where things go very wrong. All such conditions help shape a strategy.

Strategies also vary in the mechanism for using observed *N* or *n* to find the rest of the sampling plan parameters in the midst of production. That observed value might be looked up in a table of all possible sampling plans. It may be submitted to a program algorithm where if-then statements or computational formulae assign other parameter values. A well-designed if-then algorithm or formula may generate plans different from those in a look-up table, but with no effect on effectiveness and little on efficiency.

Strategies may be developed using computer programs capable of generating the *AOQL* and *IP* statistics for relevant ranges of *N*, *n*, *c*, and *p* values. The first step in the program is to generate all sampling plans for each value of *N* (or *n*) that has an *AOQL* under the error limit. The next step is to choose the plan, for each *N* (or *n*), with the minimum *IP*(*p*), where *p* is the best guess or presumed level of incoming error. A listing of the resulting plans comprises a look-up table.

Developing algorithms or formulae to replace the look-up table may have advantages, perhaps less code or risk of file corruption, in implementing a QC program. The algorithm or formula should fit the table in the sense that it generates the same plan listed in the table. Finding that alternative rule still is a partly trial-and-error effort, using statistics derived from the table. One area of development for sampling plan strategies is to find more deterministic means of fitting good formulae to a table.

A final step in developing strategies is to run a program that checks each plan's effectiveness and documents their efficiency. It also checks to be sure none of the plans in range were invalid (resulting in adaptations illustrated in Table 2), due to conditions where the hypergeometric probability is undefined, which are most common when batches are small. The final program also shows diagnostics, such as whether the *AOQL* value was set at the highest value of *p*, a sign that it is not the maximum over a large range of p. Diagnostics may also show that parameter ranges need to be trimmed to avoid invalid sampling plans.

When completely developed, a strategy is specified by:

- The error limit required,
- Ranges of possible values for $\{N, n, c, p\}$,
- The presumed level of p, and
- The mechanism for finding $\{N, n, \textit{and } c\}$: a table, a formula, or an algorithm.

The strategy should specify also what happens when observed parameters fall outside appropriate ranges. Particularly, if a batch is too small, it may be fully verified, bypassing sample verification, or joined with another batch to form one that is acceptably large.

3.5.1. Lot-Size Strategies

A lot-size strategy begins with the observed value of N. It is useful when an operation's procedures make it impossible or burdensome to set N, but not n or c. For each batch encountered in production, these strategies apply a suitable mechanism to find values for n and c, using the observed value of N, specifying the *AOQL* requirement and any range for N that is relevant. The "N" in the labels of these examples denotes that they are lot-size strategies.

- N_1 – *AOQL* < 0.01; N: 300-3000; n and c are selected for efficiency but not otherwise constrained in range; p is presumed not much different from 0.01; The mechanism to find the sampling plan is to look up N in a list and adopt the associated n and c. (See Table 1).
- N_2 – Same as N_1, except p is presumed to be 0.005 or less, sponsors wish c > 0, and a different look-up table is generated.
- N_3 – Same as N_1, except the mechanism to find a sampling plan involves computing:
  $c = 1 + \text{int}(N / 1000)$ and
  $n = \max(78, \text{ceil}(200*c/3))$.

N_3 is an example of a set of formulae that attempt to recreate the contents of a look-up table such as Table 1. The formulae do not always produce the same sampling plans found in the table. If the table has all the most efficient plans, the formula will not always yield as efficient a plan. It is important to double check that any plan generated by a formula still has an *AOQL* less than the required limit.

Table 1. Rows from a Look-up Table for Strategy N_1

N	n	c
300	68	1
. . .		
451	72	1
452	109	2
. . .		
1065	166	3
1066	122	2
1067	167	3
. . .		
3000	289	5

3.5.2. Sample-Size Strategy

A sample-size strategy is useful when an operation's procedures make it impossible or burdensome to set n, as well as N. For each batch encountered in production, these strategies apply a suitable mechanism to find a value c, using the observed value of n, checking that values of N and n suit range specifications and the *AOQL* requirement. In this strategy, a small n (rather than N) signals the need for full verification. The "n" in the labels of these examples denotes that it is a sample-size strategy.

- n_1 – *AOQL* < 0.01; N: 300-3000; n: 66-292; no presumed level of p; The mechanism to find the sampling plan is to look up n and range of observed N and adopt the associated value for c. (See Table 2).
- n_2 – same as n_1 except the mechanism to find a sampling plan involves computing:
  $c = \text{int}((0.015 * n) - 0.23)$.

Table 2. Rows from a Look-up Table for Strategy n_1

n (range)	N (range)	c
66 – 82	1501 - 3000	invalid*
66 - 82	300-1500	0
83 - 132	2431 - 3000	0
83 - 132	300 - 2430	1
133 - 188	300 - 3000	2
189 - 246	300 - 3000	3
247 – 292	300 – 3000	4

* Some plans in these ranges are invalid – resample, pool batches, verify 100% or otherwise reform the batch.

The n_2 strategy needs only one formula to completely designate a sampling plan since both N and n are observed from the batch.

Table 2 could easily be turned into an algorithm by generating a line of computer code for each row, like:
If 65< n < 83 and 299< N < 1501 then c = 0.

4. Limitations

- This paper follows the recommendation of Dodge (1963) to use the *AOQL* rather than the Acceptable Quality Level or Lot Tolerance Percent Defective acceptance sampling approaches. Those other approaches are not evaluated in this paper.
- Data about one strategy may not always be validly compared to that of another. The intent of presenting Table 3 and 4 data in this report is to illustrate the variety in strategies and related statistics. In general, strategy comparisons should be limited to alternatives applicable to a specific QC operation. For example, in selecting a strategy to implement in a given situation, you might compare the projected efficiencies of alternatives that differ only in the mechanism formula. After a strategy has been implemented on a second occasion, results may be compared meaningfully. Also, projections, such as those in Table 3, should not be treated as a prediction or standard for implementation results, such as those in Table 4.
- It is likely that algorithms or formulae presented in this paper could be improved to better fit the list of optimal plans and thus yield marginally more efficient plans overall. This is an area for future development.
- Incoming error and outgoing error percents are estimated with the assumption that errors among fields not inspected are best represented by errors among fields that fell into inspection.
- Similarly, standard errors of those percents are estimated without data transformations to compensate for error rates' proximity to zero (Wolter, 1985, p. 368), or stratification to take into account inherent batch structure (Cochran, 1977, p. 66).
- For better operational control and less operational burden, keying and coding production both implemented full rework of failed batches. That meant sample units were keyed or coded an extra time rather than skipped over when the rest of a failed batch was verified. Observed workload computations reported in this paper ignored that extra work in order to be consistent with *AOQL* computations. Actual efficiency differed as a result, but the effectiveness of the QC did not.
- Diagnostic review of 2005 QC data showed a few instances where a file management step or output record was missed in implementing QC. No more than one percent of QC data was or might have been affected in either operation.

5. Illustrating Sampling Plan Strategies

Insights into strategies for selecting sampling plans arise from observing:

- statistics projected from sampling plans representing all those that could be generated by a strategy, and
- results of recent operations applying a strategy.

Projected effectiveness and efficiency statistics illustrated here may be the best tools for judging the fit of prospective strategies to an operation. Observed statistics illustrated here are useful for evaluating one occasion's implementation of an operation against another, if they are based on the same strategy.

5.1. Projected Performance of Strategies

Statistical projections using *AOQL* and *IP* formulae illustrate how individual sampling plan strategies may be expected to perform in a general way. However, they do not predict results of a specific operation implementing the strategy, because sampling plans called up in the operation will likely come mostly from a small area within the limits of the strategy. The projected statistics in Table 3 show valid sampling plans with parameter conditions noted in the table. The set of plans generated for this analysis span uniformly those one would find within the constraints of the strategy. That is, for every possible observed N (or N, n combination) value that could be observed, the most efficient effective plan is included in this statistical summary. The strategies' mean *AOQL* and *IP* values are comparable with the understanding that those statistics were computed from a complete set (not a sample) of valid sampling plans generated given the specified error limit and those size ranges.

The first thing to notice in Table 3 is the difference in the number of plans implied by two types of strategies. Lot-size strategies give the flexibility to pick the most efficient combination of n and c so there is only one plan per N. Sample-size strategies have to provide a plan for every combination of N and n that may be observed. That does not complicate the task of finding an effective plan too much, because the N has a relatively small impact on *AOQL*, but each plan should be checked in case an increase in N requires adjusting c to keep the *AOQL* under the limit.

Table 3. Projected Efficiency of Certain QC Strategies Given Effective Limit of One Percent Error, Ranges of Parameters (*N*, *n,* and *c*), Presumed Level of *p*, and Mechanism Employed: Averages of *AOQL* and *IP* Statistics Computed Over Valid Sampling Plans with a One Percent Error Limit.

Strategy	N_1	N_2	n_1	n_2
N	300-3000	300-3000	300-3000	300-3000
n	66-292	66-82	66-292	66-292
c	0-5	1	0-4	0-4
assume *p*	~ 0.01	~ 0.005	~ 0.01	~ 0.01
Mechanism	table	table	table	formula
AOQL	0.00996	0.00986	0.00682	0.00642
IP(0.0)*	12.4	6.6	15.3	15.3
IP(0.005)	13.1	11.7	18.5	19.6
IP(0.01)	19.9	23.9	34.6	39.4
# plans	2701	2701	613127	613127

* *IP*(0.0) = 100**n/N* , since the rate of inspection equals the sample rate when there is no error.

Since the most efficient effective plans have an *AOQL* close to but less than the error limit, the proximity of the mean *AOQL* to that limit is one index of the efficiency of plans in that strategy. The *IP* means project efficiency for different specific presumed levels of incoming error. Since lot-size strategies narrow the universe of plans on the basis of efficiency, the efficiency of sample size strategies is generally less. That generalization may be less dramatic if we take into account any way of trimming the universe of sample-size strategy plans, perhaps knowing *n* is close to some given percent of *N*. That possibility is not illustrated in Table 3.

Even for lot-size strategy plans, the *AOQL* can be much less than 0.01, suppressing errors lower than the nominal specified limit. The specified limit is conservatively assured, especially with more and more batches, but it still is not guaranteed absolutely.

The *IP* statistics provide the best indices of efficiency when there is some confidence about the specific level of incoming error. An *IP* (0.005) value represents the percent of the production workload that would have to be inspected to assure no more than one percent outgoing error – as long as the unknown incoming error is actually 0.5 percent. Small *IP* values imply greater efficiency. Efficiency is better if incoming error is low, so *IP*(0.005) means are lower than *IP*(0.01) means in Table 3. If there is no incoming error, the inspection is limited to QC sample cases and the workload is at the minimum, the sampling rate. (Formula 5 is strictly undefined for $p = 0$, so *IP*(0.0) is defined directly as the sample rate.)

The value of comparing projected efficiencies for alternate strategies designed for a given operation is illustrated in Table 3. N_2 is designed for the operation based on convincing evidence that incoming error is near or below 0.5 percent, as is expected in the response coding operation, where a small experienced staff of experts define the code values and procedures as well as implement the coding. N_1 is designed for the same type of work but for much larger workloads and the influx of less skilled coders needed to do the coding. Thus the plans selected for N_2 have lower IP(0.005) and higher IP(0.01) values than N_1. If the low-error assumption is true, N_2 will be more efficient. If incoming error is close to the one percent level, N_1 will be more efficient.

5.2. Performance Data and Observations

QC results for the expert response coding of the 2005 National Census Test write-in fields illustrate a lot-size strategy similar to strategy N_2, described in section 3.5.1, but with lower batch-size range limits. QC for the keying of the 2005 National Census Test questionnaires illustrates a sample-size strategy, essentially strategy n1, described in section 3.5.2.

Results reported in Table 4 are based on counts of errors observed or projected over all batches, since such overall statistics correspond to how the data were used. The percent sampled is defined as the sum, over all sampling plans employed, of all n divided by the sum of all plans' *N*, regardless of decisions to fully verify batches for operational reasons other than an invalid sampling plan. If a sampling plan is not valid because the batch was too small, $n = N$. Percent inspected is the same as percent sampled, except *N* is substituted for *n* in batches fully verified after a batch failure.

The incoming error percent is the percent of production units determined to be in error out of all inspections done. The outgoing error percent is the estimated percent of error left in coded or keyed fields at final transmission. Such errors exist only in fields that were never inspected, since rectification means inspected units were transmitted with corrections. Specifically, the outgoing error estimate is the incoming error estimate multiplied times the portion of units transmitted that were not inspected (one minus the proportion inspected). By definition, outgoing error must be less than incoming error, since rectification can only reduce incoming error. Standard errors are provided, in parentheses, for these error estimates, since they were based on sampled data, unlike most results reported in this paper, which were computed from the full complement of available data.

Table 4. Results of 2005 QC Trials

	Coding	Keying
Strategy	Similar to N_2	Similar to n_1
Number of batches	48	4489
Number of fields	12033	7252781
Percent sampled	26.67	10.05
Percent inspected	33.82	19.11
Percent incoming error	0.12 (0.043)	0.88 (0.007)
Percent outgoing error	0.08 (0.035)	0.71 (0.006)

The data in Table 4 are presented to illustrate the kinds of metrics that can be generated in the context of applying *AOQL* methodology. Coding and keying operations are so different that there is not much point in comparing them other than to show that different operations may be served by sampling plan strategy tools.

Developers and management might use such information to better understand, control, and design QC processes. Such results could be usefully compared to those from different occasions implementing the same operation or against expected outcomes. The percents sampled and inspected are higher in coding results than is common. When batch sizes are so small (average batch size is around 250), sample verification needs to be replaced with full verification. Another strategy that better fits the operation or a non-burdensome change to procedures to increase batch size could improve the efficiency of coding QC. On the other hand, the actual estimate of incoming error for coding QC confirms prior understanding that p may be presumed low, justifying a strategy like N_2 rather than N_1.

Overall, the results of Table 4 are realistic for *AOQL* applications. In keeping with the QC requirement for both coding and keying, the outgoing error estimates are indeed less than one percent (Standard errors are reported in parentheses in Table 4.). Since the incoming error estimate is also less than one percent and outgoing error can only be less than incoming error, *AOQL* methodology is not the sole cause of that success. Data from operations with higher incoming error could demonstrate how rectification and full verification of failed batches control outgoing error.

6. Conclusions

1. The *AOQL* statistic is a tool useful for identifying one or more acceptance sampling plans that are effective in assuring a specified limit to outgoing error.

2. The $IP(p)$ statistic is a tool useful for selecting, from a set of sampling plans, the most efficient in terms of minimal inspection workload, given information or best assumption about the level of incoming error p.

3. Using those tools and information about operational constraints, strategies for selecting effective and efficient sampling plans can be developed even when lot sizes or sample sizes cannot be controlled.

7. Recommendations

1. Be aware of burden and ways to minimize it, including by adapting a sampling plan strategy.

2. Use and evaluate sampling plan strategies with these tools.

3. Research refined techniques for fitting algorithms or formulae in developing sampling plan strategies.

4. Engage operational stakeholders in setting meaningful error limit requirements and monitoring the effects of using them.

References

Cochran, W. G. (1977). *Sampling Techniques*, fourth edition, Wiley, New York, NY.

Dodge, H. F. (1963). Choosing Acceptance Sampling Plans, *Industrial Quality Control*, R. A. Freund (ed.), Vol. 20, No. 2.

Montgomery, D. C. (2001). *Introduction to Statistical Quality Control*, fourth edition, Wiley, New York, NY.

Wadsworth, H. M., Stephens, K. S., and Godfrey A. B. (2002). *Modern Methods for Quality Control and Improvement*, second edition, Wiley, New York, NY.

Wolter, K. M. (1985). *Introduction to Variance Estimation*, Springer-Verlag, New York, NY.