University of Hertfordshire UH 25 Research Archive

Citation for published version:
Neil H. Spencer, & Lindsey Kevan de Lopez, 'Item-by-item sampling for promotional purposes', *Quality Technology & Quantitative Management*, first published online 7 June 2017.

DOI:
https://doi.org/10.1080/16843703.2017.1335494

Document Version:
This is the Accepted Manuscript version.
The version in the University of Hertfordshire Research Archive may differ from the final published version. **Users should always cite the published version of record**.

Copyright and Reuse:
This manuscript version is made available under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

Enquiries
If you believe this document infringes copyright, please contact the Research & Scholarly Communications Team at rsc@herts.ac.uk

Item-by-item sampling for promotional purposes

Neil H. Spencer* and Lindsey Kevan de Lopez

Statistical Services and Consultancy Unit, Hertfordshire Business School, University of Hertfordshire, Hatfield, Hertfordshire, AL10 9AB, United Kingdom

Abstract: In this paper we present a method for sampling items that are checked on a pass/fail basis, with a view to a statement being made about the success/failure rate for the purposes of promoting an organisation's product/service to potential clients/customers. Attention is paid to the appropriate use of statistical phrases for the statements and this leads to the use of Bayesian credible intervals, thus it exceeds what can achieved with standard acceptance sampling techniques. The hypergeometric distribution is used to calculate successive stopping rules so that the resources used for sampling can be minimised. Extensions to the sampling procedure are considered to allow the potential for stronger and weaker statements to be made as sampling progresses. The relationship between the true error rate and the probabilities of making correct statements is discussed.

Keywords: Bayesian credible intervals, hypergeometric distribution, marketing, sampling.

1. Background

Undertaking hypothesis testing in a sequential manner, building up a sample until sufficient data have been collected, has much of its origins in the work of Wald [14]. Often used in a quality control context, these methods are designed to minimise the number of items examined before a

* Corresponding author. Email: N.H.Spencer@herts.ac.uk

decision is made. This can be particularly important in the context of inspections which destroy the item concerned or make it unusable or unsellable in order for a judgement to be made as to whether it is of acceptable quality or not. Outside quality control, sequential procedures have also been used in a variety of other settings such as in clinical trials to minimise the number of patients that need to undergo arduous or potentially dangerous procedures or medication (e.g. [15]), in animal experimentation to minimise the number of subjects that might experience degrees of suffering (e.g. [6]), in psychology to minimise the number of people that are needed for the successful assessment of complex interventions (e.g. [7]), in auditing to minimise the sample required to achieve a desired precision (e.g. [9]). In all these examples, sequential sampling methodologies are being used in order to obtain good sample data at minimum economic or social cost without compromising the subsequent statistical analysis. There are a number of books devoted to the subject (e.g. [11, 13]).

Since Wald's introduction of the sequential probability ratio test (SPRT), further methodological developments have taken place to extend its use and provide for a variety of different situations where sequential analysis is advantageous. These range from developments of methods for examining decision making in psychology (e.g. [1, 12]) to advances for its use in biological applications (e.g. [5, 8]) to progress in clinical trials (e.g. [3, 16]) to adaptations for business and industry (e.g. [2, 4]). The extent is so vast that it is not practical to give a reasonable summary here. However, there are two reasons why existing methodologies are not applicable to the situation given in this paper. Firstly, most sequential analysis techniques assume that the items tested come from a large (potentially infinite) population (relative to the number of items tested) and can thus use approximations that take advantage of this. Here, we have a finite population and the sample tested may be a sizeable proportion of this, meaning that it would not be appropriate to

use these approximations. Secondly, the analysis must produce a Bayesian credible interval rather than a hypothesis test or confidence interval which are typically the result of a sequential analysis.

The motivation for this paper is the scenario where an organisation wishes to obtain an estimate of the true proportion of unacceptable items in a population and use this for promotional purposes (and other objectives). That is, it wishes to promote its product or services by showing potential clients/customers that the probability of unacceptable items is very low. In this paper we use the example of identifying incorrect records in a database but equally valid scenarios could be faulty products from a manufacturing process, incorrect diagnoses from medical images, substandard customer interactions in a call centre, etc.

In section 2, we discuss the statistical measure that is relevant to the probability statement required. In section 3 we address the methodology and in section 4 we demonstrate an implementation of the methodology. The issue of prior probabilities is discussed in section 5 and extensions to the sampling procedure are discussed in section 6. The issue of there being multiple stopping points (and hence multiple opportunities to make a false statement) is dealt with in section 7 and conclusions are given in section 8.

2. Choice of statistical measure

A natural form of wording for a statement that an organisation may want to make for promotional purposes (while retaining mathematical rigour) is “There is a 95% probability that the true proportion of incorrect records is less than 5%”. Neither a hypothesis testing approach nor a confidence interval can provide a statement of this type but this can be accomplished with a one-sided Bayesian credible interval. It is thus a form of analysis which can yield such an interval that is employed in this paper.

The issue of monetary cost is sometimes included in sequential sampling methodologies. This can be the case when cost can be easily expressed in monetary terms but in many situations, costs are not so easily represented. This is the case for the examples of clinical trials and animal experimentation given in section 1 and is also true for promotional purposes where the cost of making an incorrect claim is in terms of reputation. As a result, methods which require an expression of monetary cost cannot be used here.

3. Methodology

We define the size of the population to be *N* records, the unknown number of incorrect records in the population to be *R*, the size of the sample examined to be *n* records and the number of incorrect records in the sample to be *r*.

For a desired maximum error of R^*/N and the requirement to make a statement that there is at least a $100\times(1-\alpha)\%$ probability of the true error rate being no more than this, we want to find *R** such that the following probability statement is true where the probability of the number of incorrect records in the population (*R*) is conditional upon the number of incorrect records in the sample (*r*), taking into account the sample size (*n*) and the population size (*N*).

$$\Pr\left(R \le R^* \middle| r; n, N\right) = \sum_{R=0}^{R^*} \Pr\left(R \middle| r; n, N\right) \ge (1-\alpha)$$

As $\Pr\left(r \middle| R; n, N\right) = \dfrac{\Pr\left(R, r; n, N\right)}{\Pr\left(R; n, N\right)}$, we get the following:

$$\Pr\left(R \middle| r; n, N\right) = \frac{\Pr\left(R, r; n, N\right)}{\Pr\left(r; n, N\right)} = \frac{\Pr\left(r \middle| R; n, N\right)\Pr\left(R; n, N\right)}{\sum_{R'=r}^{N-(n-r)} \Pr\left(r \middle| R'; n, N\right)\Pr\left(R'; n, N\right)}.$$

Under an assumption of equal prior probabilities (examined further below), we have $\Pr(R';n,N)=\Pr(R;n,N)=p^*$ for all R, R' (i.e. the probability of R incorrect records in the population is constant for all values of R). This allows a simplification of the above formula, giving

$$\Pr(R|r;n,N)=\frac{\Pr(r|R;n,N)p^*}{p^*\sum_{R'=r}^{N-(n-r)}\Pr(r|R';n,N)}=\frac{\Pr(r|R;n,N)}{\sum_{R'=r}^{N-(n-r)}\Pr(r|R';n,N)}.$$

In the denominator of this fraction we have the sum of probabilities of observing r incorrect records in the sample under different conditions: conditional on the number of incorrect records in the population (R'), given the sample size (n) and the population size (N). We are operating under a system governed by the hypergeometric distribution and it can be shown empirically that

$$\sum_{R'=r}^{N-(n-r)}\Pr(r|R';n,N)=\frac{N+1}{n+1}.$$

Hence $\Pr(R|r;n,N)=\frac{n+1}{N+1}\Pr(r|R;n,N)$ and thus

$$\Pr(R\le R^*|r;n,N)=\sum_{R=0}^{R^*}\frac{n+1}{N+1}\Pr(r|R;n,N)=\frac{n+1}{N+1}\sum_{R=0}^{R^*}\Pr(r|R;n,N)=\frac{n+1}{N+1}\sum_{R=r}^{R^*}\Pr(r|R;n,N)$$ as

$\Pr(r|R;n,N)=0$ for $R<r$.

Thus, for r = 0, 1, 2, ..., we want to identify n_0, n_1, n_2, … which are the smallest n such that

$\frac{n+1}{N+1}\sum_{R=r}^{R^*}\Pr(r|R;n,N)\ge(1-\alpha)$ or equivalently $\sum_{R=r}^{R^*}\Pr(r|R;n,N)\ge(1-\alpha)\frac{N+1}{n+1}$.

The above calculations for the Bayesian credible interval are such that the required sample size cannot be known before sampling begins. One cannot know how many incorrect records will have

been identified after a particular number of cases have been sampled and thus the required sample size will have to be monitored as sampling proceeds and updated whenever an incorrect case is found. As an incorrect record could be found at any point in the sampling, monitoring must be on a case by case basis and sequential sampling, a form of acceptance sampling (see, for example, [10]) is thus being undertaken.

In acceptance sampling, items from the batch are chosen randomly and sequentially on an individual basis and either judged to be faulty or not. If no faulty items are identified then the sampling will stop after a prescribed number of items have been investigated and the batch will be declared to be of acceptable quality. If one or more faulty items are identified then the sampling continues. Eventually the number of faulty items relative to the number of items inspected will trigger a stopping criteria and the batch accepted or rejected. Alternatively the number of samples inspected will reach a predetermined maximum and a default decision made to accept or reject the batch.

There are two reasons why the acceptance sampling approach is not appropriate in this paper. The first concerns the construction of the stopping rule which, in acceptance sampling, comes from the Sequential Probability Ratio Test (SPRT) developed in [14]. This is based on the binomial distribution which is a good approximation to the hypergeometric distribution when N is large relative to n. However, in the scenario described in this paper, it cannot be assumed that N will be large relative to n, and thus we cannot use the method of [14].

The second reason returns us to the rigour with which we wish to make a probability statement about the population. Above, we have shown that using the Bayesian credible interval allows us to make the probability statement desired. By contrast, acceptance sampling methods are not

designed for this purpose. They are used with the intention of deciding whether or not a batch of items is of sufficient quality to be accepted and thus the question they are designed to answer is not the same as required here. As the main motivation of this paper is to use a methodology which is precisely aligned with the wording of the probability statement to be used, it is clear that despite parallels existing, these acceptance sampling techniques are not appropriate here.

4. Implementation

Above, it has been shown that to implement the sample strategy required to form the probability statement, for $r = 0, 1, 2, ...$, it is necessary to identify $n_0, n_1, n_2, \ldots$ which are the smallest n such that $\frac{n+1}{N+1}\sum_{R=r}^{R^*}\Pr(r|R;n,N) \geq (1-\alpha)$ or equivalently $\sum_{R=r}^{R^*}\Pr(r|R;n,N) \geq (1-\alpha)\frac{N+1}{n+1}$.

Once sampling has commenced, if no incorrect records are found by the time n_0 records have been inspected, the sampling will stop. If one incorrect record has been found after n_0 records have been inspected then the sampling will continue until a total of n_1 records have been inspected and, if no further incorrect records have been found, it will stop. This pattern could potentially continue until all N records in the population have been examined. However, it is probable that the stopping strategy would be adapted, as discussed in section 6, before this point.

For example, if the requirement is for the statement "There is a 95% probability that the true proportion of incorrect records is less than 5%", we have $\alpha = 0.05$ and $R^*/N = 0.05$. If $N = 300$ records then $R^* = 15$. Applying the above sample size calculations yields Table 1. It shows the minimum number of records that must be checked and if no incorrect records are discovered, the sampling stops and the statement is valid. If one incorrect record has been identified in the initial sample of 50, we see from the table that the sample size required has increased to 77. If this point

is reached with no more incorrect records being identified then the sampling stops and the statement is valid. The procedure is additionally shown in graphical form in Figure 1. In this figure, if the combination of sample size and number of incorrect records reaches the line (including the part of the line running along the horizontal axis) then the sampling stops and the statement is valid.

[TABLE 1 HERE]

[FIGURE 1 HERE]

5. Prior probabilities

In sections 3 and 4, the calculations make an assumption of equal prior probabilities such that $\Pr(R';n,N)=\Pr(R;n,N)=p^*$ for all R, R' where R, R' are alternative numbers of incorrect records in the population, N is the size of the population and n is the sample size.

An alternative assumption for the prior probabilities could be that smaller values of R are more probable than larger numbers. However, it is not unreasonable to suppose that there might be some systematic error in the record entry, due perhaps to an incorrect application of a recoding algorithm, inconsistency in accuracy over time or different individuals entering the data. In these circumstances, any value of R is now feasible and it is difficult to justify any particular assumption being made about the prior probabilities. However, the equal prior probabilities assumption can be seen as adopting a neutral position and is thus appropriate in this work.

6. Extension to sampling procedure

While the desired statement might initially be “There is a 95% probability that the true proportion of incorrect records is less than 5%”, clearly a preferable statement is “There is a 95% probability

that the true proportion of incorrect records is less than 2%" (or some other figure smaller than 5%). Additionally, if the number of incorrect records is such that the statement using 5% is not possible then a weaker statement using 10% may be possible.

For each of the 2% and 10% statements, tables such as Table 1 can be constructed using the calculations of section 3. Figure 1 can be adapted to include additional lines for each alternative statement and this is shown in Figure 2. Whilst 50 is the minimum sample size needed to be able to make the 5% statement, if this position is reached with no incorrect records identified, the sampling could continue. If a sample size of 103 is reached with still no incorrect records then the 2% statement can be made. However, if after proceeding beyond 50 an incorrect record is discovered before 77 records have been checked then the trace on the graph will go above the 5% statement line and sampling would have to continue until the trace again reaches it. Alternatively, if several incorrect records are found at a relatively early stage of the sampling, it may be decided that the 10% statement will be used or no statement at all will be made.

[FIGURE 2 HERE]

7. Multiple stopping points

We have a series of potential stopping points and thus multiple opportunities for sampling will stop and a statement about the accuracy of the recording made. This is akin to the multiple testing problem in hypothesis testing when the probability of a Type II error occurring becomes inflated.

In practice, it is likely that before any sampling takes place, a decision will have been made as to the maximum number of records that are to be sampled. For the procedure illustrated in Figure 2 this may well be a maximum of 103 records. This will enable the 2% statement to be made if there are no incorrect records identified but also allow the 5% statement to be made if there are 1 or 2

incorrect records. If there are between 3 and 6 incorrect records identified, the 10% statement can be made. However, sampling may stop before 103 records have been investigated. The possible stopping points are outlined in Table 2. The bracketed phrases show repetitions of previous stopping criteria.

[TABLE 2 HERE]

The probability of making a statement can be calculated for different possible values of the true error rate, assuming the stopping points in Table 2. This is shown in graphical form in Figure 3 for the probability of making a statement at the 5% level or better. A similar graph could be constructed for the probability of making a statement at the 10% level or better.

[FIGURE 3 HERE]

The probability of making an incorrect statement for the circumstances described in Table 2 is shown to be small in Figure 3. For those true error rates greater than 5%, the probabilities of making a 5% statement are less than 10%. Penalties for incorrectly making a statement in the scenario described are not great. Indeed, the greater risk is that of not making a statement when the truth is that it could be made. For true error rates near to 5%, the probability of correctly deciding to make a statement are small. This is inevitably the case with these relatively small population and sample sizes.

In other circumstances, the probability of correctly making a statement when the true error rate is less than 5% may be higher. However, at the same time, if the true error rate is above 5%, the probability of incorrectly making a statement will be higher. If this is judged to be a problem then the number of stopping points can be reduced. This can be done by removing the bracketed

stopping points in Table 2. Then, for instance, rather than stopping if one incorrect record is identified at any of the sample sizes 78 to 100 inclusive, sampling would continue until sample size 101. This provides the opportunity for more than two incorrect records to be spotted in the interval 78 to 101 and thus the 5% statement could not be made.

8. Conclusions

In this paper we have described a sampling process that enables statements to be made about the accuracy of records in a database. Emphasis has been placed upon the validity of the wording of the statistical statements made in the statement and the defensibility of assumptions. We have also identified a variety of other situations in which a Bayesian credible interval is a more useful tool than the confidence interval that is produced from a random sampling procedure. The implementation of the procedure has been described and the issue of multiple stopping points meaning an increased probability incorrect statements being made has been addressed.

References

1. Bhatia, S. (2014). Sequential sampling and paradoxes of risky choice. *Psychonomic Bulletin & Review*, 21, 5, 1095-1111.

2. Burnetas, A. & Kanavetas, O. (2012). Adaptive policies for sequential sampling under incomplete information and a cost constraint. *Applications of Mathematics and Informatics in Military Science*, 71, 97-112.

3. Chambaz, A. & van der Laan, M.J. (2014). Inference in targeted group-sequential covariate-adjusted randomized clinical trials. *Scandinavian Journal of Statistics*, 41, 104-140.

4. Chick, S.E. & Frazier, P. (2012). Sequential sampling with economics of selection procedures. *Management Science*, 58, 3, 550-569.

5. Elliott, N.C., Brewer, N.J., Giles, K.L., Backoulou, G.F., McCornack, B.P., Pendleton, B.B. & Royer, T.A. (2014). Sequential sampling for panicle caterpillars (lepidoptera: noctuidae) in sorghum. *Journal of Economic Entomology*, 107, 2, 846-853.

6. Fitts, D. A. (2011). Minimizing animal numbers: The variable-criteria sequential stopping rule. *Comparative Medicine*, 61, 206-218.

7. Frick, R. W. (1998). A better stopping rule for conventional statistical tests. *Behavior Research Methods, Instruments, & Computers*, 30, 690-697.

8. Heath, C.A.E., Main, D.C.J., Mullan, S., Haskell, M.J. & Browne, W.J. (2015). Sequential sampling: a novel method in farm animal welfare assessment. *Animal*, 10, 2, 349-356.

9. Hitzig, N. B. (1998). Detecting and estimating misstatement in a two-step sequential sampling with probability proportional to size. *Auditing: A Journal of Practice and Theory*, 17, 1, 54-68.

10. Montgomery, D. C. (2009). *Statistical Quality Control: A Modern Introduction*. 6th Edition. John Wiley & Sons, Inc., Hoboken, New Jersey.

11. Mukhopadhyay, N. and de Silva, B.M. (2009). *Sequential Methods and their Applications*. Chapman & Hall/CRC, Boca Raton, Florida.

12 Ratcliff, R. & Smith, P.L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, 111, 2, 333-367.

13. Siegmund, D. (1985). *Sequential analysis: Tests and confidence intervals*. Springer, New York.

14. Wald, A. (1947). *Sequential Analysis*. John Wiley & Sons, New York.

15. Whitehead, J. (1997). *The Design and Analysis of Sequential Clinical Trials*. 2nd Edition. John Wiley & Sons, Ltd., Chichester, U.K.

16. Zhao, S., Cook, A., Jackson, L. & Nelson, J. (2012). Statistical performance of group sequential methods for observational post-licensure medical product safety surveillance: a simulation study. *Statistics and Its Interface*, 5, 381-390.

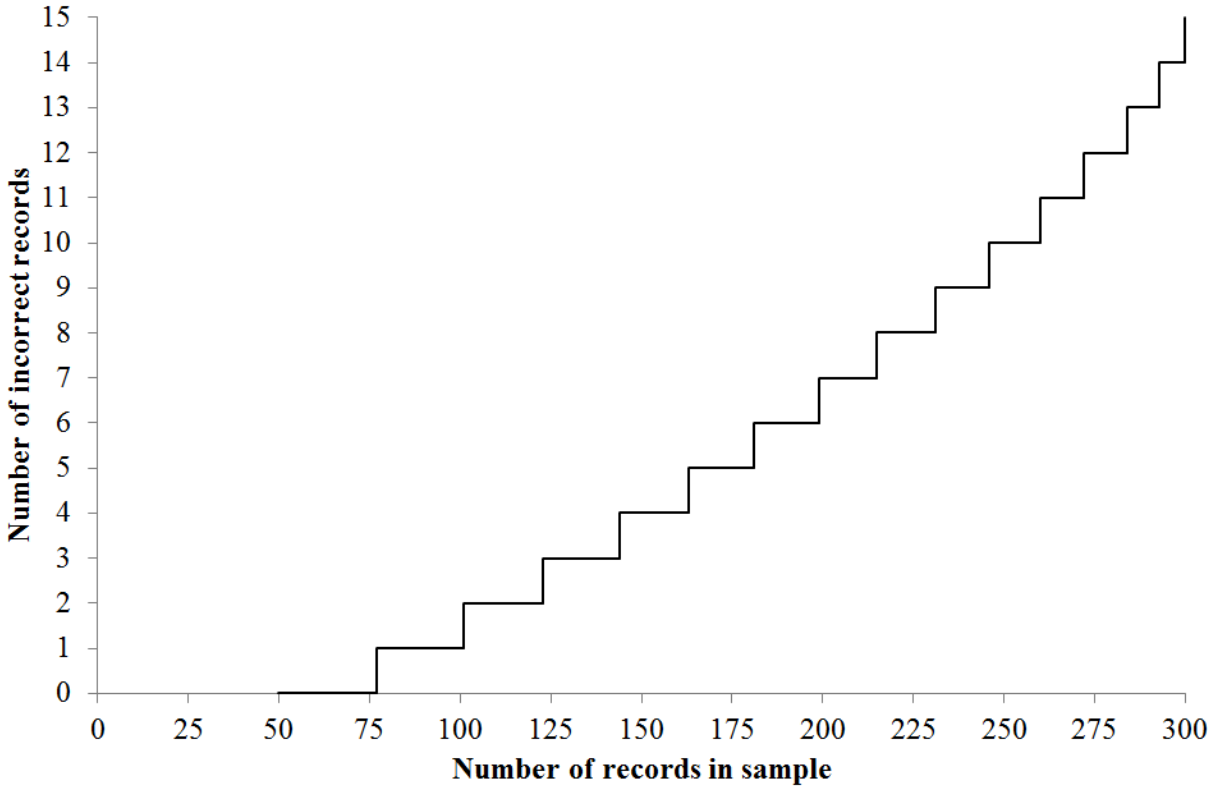

Figure 1. Stopping criteria for sampling scheme (5% statement).

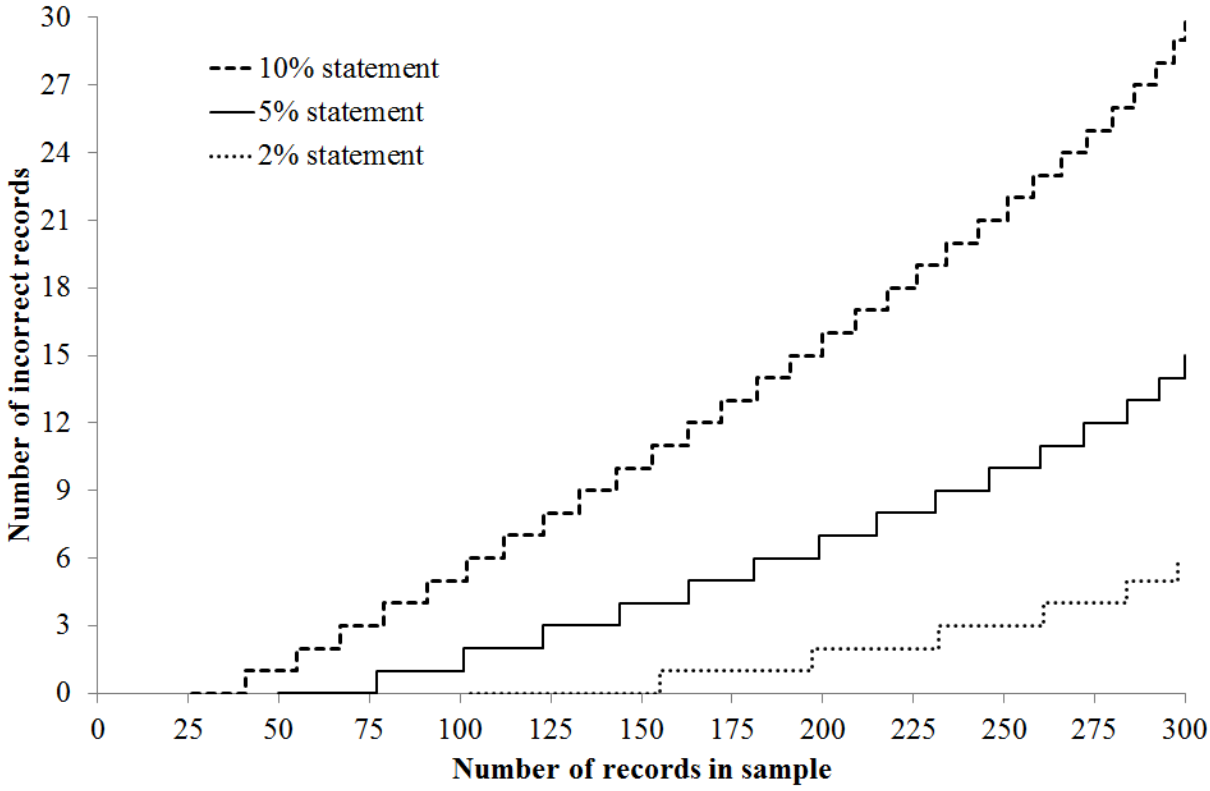

Figure 2. Stopping criteria for sampling scheme (10%, 5% and 2% statements).

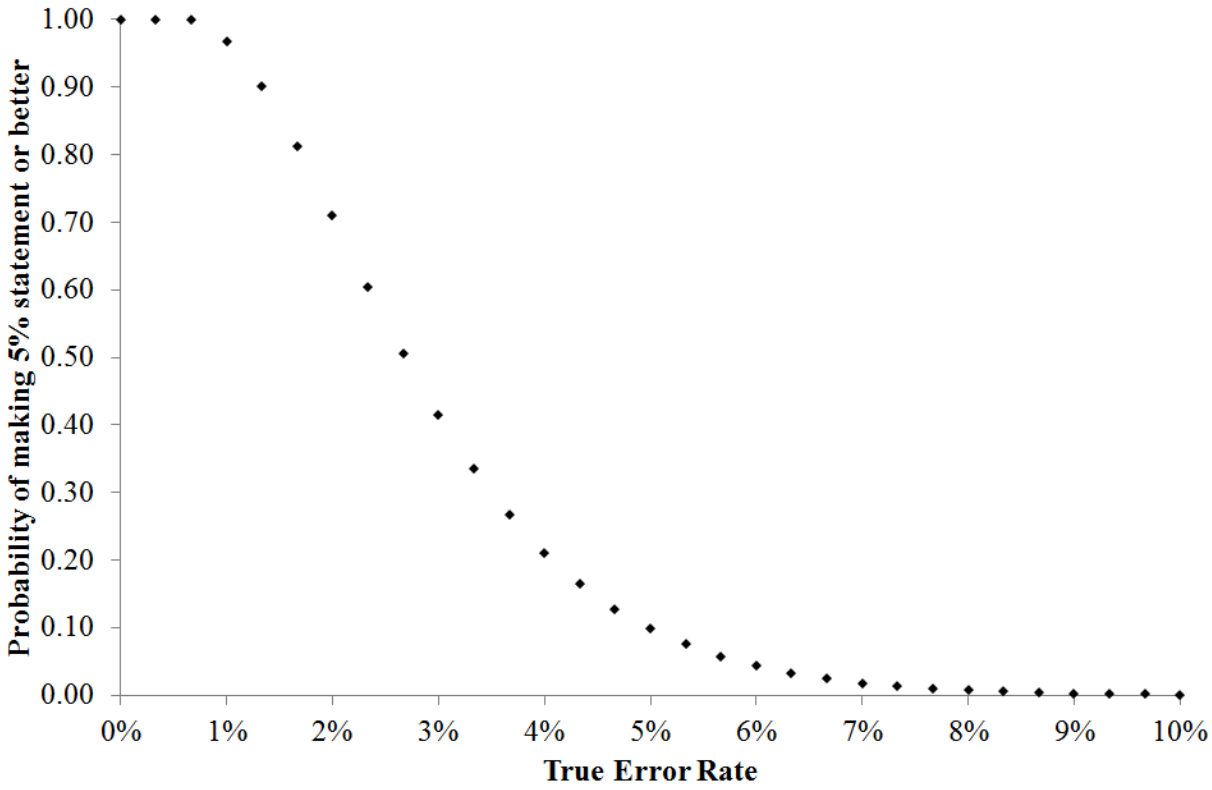

Figure 3. Probability of making 5% statement (or better) for different true error rates.

Table 1. Sample size requirements for varying numbers of incorrect records in sample (5% statement).

Number of incorrect records	Minimum sample size for number of incorrect records for 5% statement	Number of incorrect records	Minimum sample size for number of incorrect records for 5% statement
0	50	8	215
1	77	9	231
2	101	10	246
3	123	11	260
4	144	12	272
5	163	13	284
6	181	14	293
7	199	15	300

Table 2. Stopping points with maximum sample size of 103.

Sample size	Stop and make 2% statement if:	Stop and make 5% statement if:	Stop and make 10% statement if:	Stop and make no statement if:
0 to 66				more than 6 incorrect
67			3 incorrect	
68 to 76			3 incorrect	
77		1 incorrect	3 incorrect	
78		(1 incorrect)	3 incorrect	
79		(1 incorrect)	3 or 4 incorrect	
80 to 90		(1 incorrect)	(3 or 4 incorrect)	
91		(1 incorrect)	3, 4 or 5 incorrect	
92 to 100		(1 incorrect)	(3, 4 or 5 incorrect)	
101		1 or 2 incorrect	(3, 4 or 5 incorrect)	
102		1 incorrect*	3, 4, 5 or 6 incorrect	
103	none incorrect	1 incorrect*		

* Would also stop here if 2 were incorrect but if this were the case then would have already stopped at sample size defined in previous line.