Distribution of Aflatoxin in Pistachios. 7. Sequential Sampling

Thomas F. Schatzki†

Western Regional Research Center, Agricultural Research Service, U.S. Department of Agriculture, Albany, California 94710

Sequential sampling for aflatoxin testing in pistachios is evaluated using the aflatoxin distribution and Monte Carlo results previously obtained (*J. Agric. Food Chem.* **1999**, *47*, 3771−3775). The sequential protocol is modeled on the current EU test protocol by applying a three-step sampling, using 10, 20, and 30 kg sample averages. An acceptance level of 15 ng/g of total aflatoxin, under consideration for U.S. standards, is applied. Optimization leads to indifference regions of 2−30 ng/g for the first two steps. The resulting OC curve approximates that for a single 50 kg sample. The sequential protocol is applied to the results for a set of 1293 lots of the 1998 crop year, each tested with a single 10 kg sample. Ninety-five percent of the lots would have been accepted on the basis of the single test and 1.5% would have been rejected, whereas 3.5% of the lots would have required retesting.

Keywords: *Aflatoxin; pistachios; sequential sampling; sample size reduction*

INTRODUCTION

Aflatoxin, a potent carcinogenic mycotoxin, may contaminate a number of granular foods, including tree and ground nuts and grains. In many of these foods, and in particular in tree nuts, preharvest contamination will occur in individual nuts, without any evidence for internut contamination. As a direct result, the overall aflatoxin of a lot is carried by very few nuts (as few as $1/10^5$). The contamination level among the contaminated nuts varies widely, extending over 8 orders of magnitude ($0.01-10^6$ ng/g). The resulting aflatoxin distribution is thus exceedingly broad and skewed.

The sampling statistics resulting from such a distribution have been analyzed in some detail, particularly for pistachios (Schatzki, 1995a,b, 1998, 1999). The governing equation predicts that the variance is given approximately by 800000 × *m*/number of nuts tested, where *m* is the lot mean. A value of 700 nuts/kg (20 nuts/once) will be used. Extremely large samples (or the average of a large number of small samples) are required to obtain a reliable estimate of the average aflatoxin level of a lot. On the basis of the above equation, to obtain a variance equal to 10% of *m*, a sample of ∼12000 kg would be required. Such a sample size would obviously be totally impractical as pistachio nuts currently cost ∼$3/kg wholesale and the entire sample would need to be destroyed. Analysis and mixing costs would be additional. As a compromise, the European Union (EU) recently adopted a sampling protocol calling for three samples of 10 kg each, which is still not negligible, but acceptable. For a 10 kg size the variance amounts to 114 × *m*, yielding a standard deviation (variance$^{1/2}$) of ∼15 ng/g at the EU acceptance level of $m = 2$ ng/g for in-shell pistachios. In fact, the aflatoxin level of shipments from a given shipper, or from a given country, can be estimated only after repeated receipts and analyses from the same source over a year. The averaging over shippers from a given country for a particular year was discussed and justified in Schatzki (1995b).

Currently, consideration is being given to setting up a sampling protocol to monitor domestic shipments in the United States (as well as some exports). An acceptance level of 15 or 20 ng/g is proposed. It would simplify the process if the sampling for domestic shipments were to match that for EU shipments, although the actual acceptance levels would differ. For the present discussion, an acceptance level of 15 ng/g is assumed. The use of a single 30 kg sample average might be a convenient solution. The 30 kg sample matches the three 10 kg samples used in the EU protocol for acceptance of lots destined for human consumption. Furthermore, a 30 kg (66 lb) sample is close to the 50 lb sample currently being used in the United States for sampling of imported pistachio lots. It is also close to the 48 lb sample used for peanuts (Whitaker and Dickens, 1989). However, a single 30 kg sample average at 15 ng/g would still yield a standard deviation of 25 ng/g. More specifically, it would still yield a broad operational characteristic (OC) curve. An OC curve can be expressed as a plot of $p(\text{acceptance}|m)$ versus *m*, where $p(\text{acceptance}|m)$ is the probability that a lot of aflatoxin average level *m* will be accepted (given a preassigned sampling protocol and acceptance level). In general, an OC curve is a sigmoidal curve, high for *m* less than the acceptance level and low for *m* larger. As long as $p(\text{acceptance}|m) < 1$ for *m* less than the acceptance level, one rejects good lots; similarly, if $p(\text{acceptance}|m) > 0$ for *m* greater than the acceptance level, one accepts bad lots, simply on statistical grounds. Hence, the steeper the OC curve, the better the test. An OC curve can be steepened by using a larger sample size. For reasons discussed above, this is not practical.

Another approach to steeper OC curves is to use sequential sampling. In its simplest form, rather than basing acceptance on a single sample of preset size (here 30 kg), one takes smaller, incremental samples, keeping track of the average test level for the samples tested so far. If this average falls below a fixed acceptance value

† E-mail tom@pw.usda.gov

10.1021/jf000342o
Published on Web 08/08/2000

(selected before the testing is started and depending on the number of samples tested so far), one accepts the lot. If the average falls above a similar (higher) rejection value, one rejects the lot. If it falls between, called the indifference region, one continues testing. Commonly, after a relatively large number of small samples, one chooses the (lower) acceptance value and the (higher) rejection value as the same number, eliminating the indifference region and forcing an accept/reject decision at that point. The number of small samples is frequently (although not necessarily) chosen so that the total weight of the smaller samples equals that of the single sample of the original test which is being replaced. In this way one ensures that in no event will more sample material be required than would have been in the single-sample case, although the amount will often be less because of an earlier decision. The number of tests may be larger, of course. By careful choice of the sequence of acceptance and rejection limits in the sequential test one may obtain a test cost well below that in a single test, on average.

This is the approach chosen here. The sequential sampling approach has been described here in some detail because the 30 kg test is ideally suited to sequential sampling. As noted, EU tests require three 10 kg samples, so these three samples can be used in a sequential test. In the case of a U.S.-bound sample, sequential testing will require additional analysis of at most two samples, but again no additional sample cost.

MATERIALS AND METHODS

Sequential analysis has been discussed in considerable detail (Wald, 1947). This author considers a number of possible protocols, but generally from a point of view different from that adopted here. In Wald, a sequential analysis plan is sought that will produce a predetermined OC curve. In the present case, because of the constraints described above, the choice of sampling plan becomes a four-parameter problem. Before we begin analysis, we choose a set of two bounds, C_1^- and C_1^+. If the first sample concentration $C_1 \le C_1^-$, we accept the lot; if $C_1 > C_1^+$, we reject it. If $C_1^- < C_1 \le C_1^+$, called the region of indifference, we take a second sample and average these two sample results to obtain C_2. We then proceed as with C_1 except now with limits C_2^- and C_2^+, which may differ from C_1^- and C_1^+. C_1^-, C_1^+, C_2^-, and C_2^+ form the four parameters. In this notation, the superscript on C indicates whether an acceptance or rejection bound is in use and the subscript indicates the number of small samples that is being averaged. If we again fall into the indifference region, we take a third sample, average all three to C_3, and accept or reject, depending on whether $C_3 \le 15$ ng/g or > 15 ng/g. The situation is shown schematically in Figure 1. There are six possible outcomes,

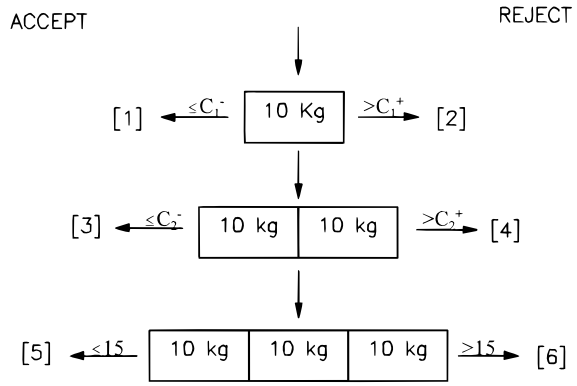

Figure 1. Sequential sampling protocol. The numbered paths correspond to the paths defined in connection with eq 1.

three corresponding to acceptance and three to rejection. The probabilities defining the acceptance paths are given by

$$[1] = p(C_1 \le C_1^-)$$

$$[3] = p(C_1 > C_1^-) \times p(C_1 \le C_1^+) \times p(C_2 \le C_2^-) \qquad (1)$$

$$[5] = p(C_1 > C_1^-) \times p(C_1 \le C_1^+) \times p(C_2 > C_2^-) \times p(C_2 \le C_2^+)) \times p(C_3 \le 15 \text{ ng/g})$$

with similar expressions for [2], [4], and [6]. The probabilities are conditional on the lot mean, m, which is left out of the expressions for simplicity. Even so, the expressions are complex but are easily derived from the probabilities by considering the arrows of Figure 1. The probability of acceptance, $p(\text{acceptance}|m)$, is given by [1] + [3] + [5]. The conditional probabilities involving C_1 and C_3 (10 and 30 kg) were directly available from Monte Carlo calculations done previously (Schatzki, 1999); the calculations for C_2 (20 kg) were computed in the same manner. (In Monte Carlo one produces a large number of nut mixtures at random according to a distribution, computes the significant results of each choice, and averages these. In effect, one simulates the statistics of sampling by computer.) As indicated, the OC curve is then simply computed from the above equations and the four adjustable parameters C_1^-, C_1^+, C_2^-, and C_2^+. What is done here is very similar to what was done by Whitaker and Dickens (1989) for peanuts except that these authors did not optimize the regions of indifference.

RESULTS AND DISCUSSION

Optimization. Such a four-parameter problem can be optimized by use of response surface methodology (RSM), although the computations would involve the tie-in of the Monte Carlo results, not a trivial problem. In RSM one computes the value of an outcome (response surface) at a large number of points in a multidimensional space, here of C_1^-, C_1^+, C_2^-, and C_2^+. One then seeks an optimal solution, typically a minimum or maximum, by tracking this surface, using methods of maximum gradients and the like. A specific response surface, measuring the desirability of the OC curve obtained, would be required. One might use the sum of false positives and negatives, that is, the area above the OC curve for $m < 15$ ng/g plus the area below for $m > 15$ ng/g. Alternatively, the slope at $m = 15$ ng/g might be used. At best, such an approach is computationally very expensive. Furthermore, as discussed below, the OC curve alone is not sufficient, however, to establish the desirability of sequential sampling. For this, as well as computational reasons, it was thought to be more desirable to study and optimize the four-parameter problem analytically.

The limits of analysis can be written immediately. For very small indifference regions, that is, $C_1^- = C_1^+ = C_2^- = C_2^+ = 15$ ng/g, one would simply have a single 10 kg sample OC curve, which is very broad. For a very wide indifference region, $C_1^- = C_2^- < 0$ ng/g (i.e., no lot is accepted on the basis of the first two small sample averages, no matter what they are) and $C_1^+ = C_2^+ \gg 15$ ng/g. In this case one would not make decisions until three samples were analyzed and one would have an OC curve corresponding to a 30 kg sample, which is still quite broad. Some single-sample OC curves are shown in Figure 2. For intermediate values of the indifference regions, particularly where the parameters fall into ranges of m where $p(C_1|m)$ and $p(C_2|m)$ change rapidly,

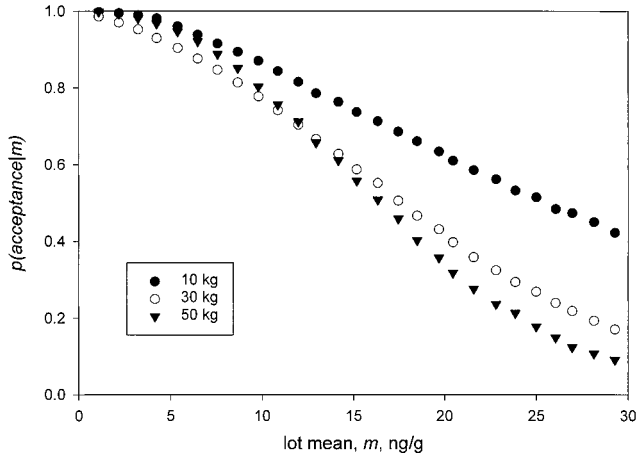

Figure 2. OC curves, in-shell pistachios, single samples. Each point corresponds to a specific value of m for which a Monte Carlo calculation was performed.

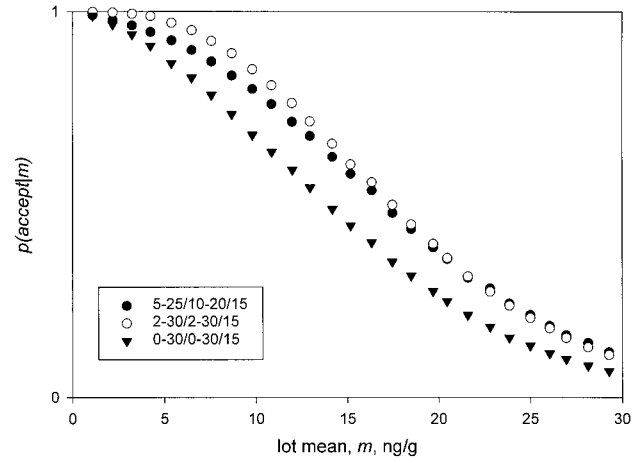

Figure 3. OC curves, in-shell pistachios, sequential samples, 10 kg each. Each point corresponds to a specific value of m for which a Monte Carlo calculation was performed. Nomenclature describing the curves refers to the sample average limits for one, two, and three 10 kg samples: $C_1^- - C_1^+/C_2^- - C_2^+/C_3^- = C_3^+$.

the resulting OC curves are more sigmoidal and thus more desirable. Broad optima of steepness are found at $C_1^- = C_2^- = 2$ ng/g and $C_1^+ = C_2^+ = 30$ ng/g. There is little or no advantage in choosing different indifference regions for one or two sample averages (10 or 20 kg). It is also found from calculations, not shown here, that little interaction occurs between the parameters. For reasons discussed below, our main interest in the OC curve and the main advantage of sequential sampling comes from increasing the OC curve at small m. Some of the more illustrative OC curves for sequential sampling are shown in Figure 3 to the same scale as Figure 2. The improvement of the OC curves, due to sequential sampling, is significant, although not large. (More severe sharpening would be achieved were the probability curves more peaked, but this would require larger samples, which was not an option here.) By overlaying Figures 2 and 3 it is found that the optimum (2−30/2−30/15) OC curve closely approximates the single 50 kg sample curve in shape, except that the sequential sampling curve is ∼0.03−0.04 higher for almost all m. That is a distinct advantage at low m.

Priors. The OC curves are interesting on their own account, but insufficient if one wishes to evaluate the effect of sequential analysis on pistachio testing. For this one needs to introduce the prior probabilities, $p(m)$, of aflatoxin levels among the lots presented for testing, as was done in Schatzki (1999). In the present case an approximation to a prior distribution is available, thanks to data from DFA of California. DFA is a not-for-profit quality control laboratory that tests pistachios on request, among other products. For the 1998 crop 1293 tests for aflatoxin were carried out, each on a 10 kg submitted sample. Data were reported to the nearest 0.1 ng/g and are summarized in Table 1 (M. Hurley, DFA of California, personal communication, 2000). Assuming the test data represent the actual average aflatoxin distribution among the lots, one has an approximation for $p(m)$. (Note that no correction was applied here for producer or, more specifically, for the amount of product produced by this producer, to obtain an overall product distribution, as was done for almonds in 1993.)

Table 1. Test Results on 1293 Lots from 1998 Pistachio Crop

range of test data, in-shell pistachios, ng/g	fraction of samples	range of test data, in-shell pistachios, ng/g	fraction of samples
0	0.902	10.1−15	0.005
0.1−0.3	0.007	15.1−20	0.002
0.4−1	0.027	20.1−31.6	0.009
1.1−2	0.014	31.7−100	0.004
2.1−3.1	0.008	100.1−316	0.001
3.2−5	0.009	317−1000	0.001
5.1−10	0.014		

The $p(m)$ assumed to be represented by Table 1 may now be convoluted (multiplied) by the results of the previous section. One obtains an estimate of the effect of sequential sampling by use of the usual Bayesian expression $p(\text{acceptance}) = \sum_m p(\text{acceptance}|m) \times p(m)$. In the evaluation of this expression the probability over a range of m, as given by the Table 1, will be used. Because $p(\text{acceptance}|m)$ is only available at specific m (from the Monte Carlo calculation described above), interpolation was used. When this calculation was carried out for a single sample of 30 kg [by computing $\sum_{m<15} (1 - p(\text{acceptance}|m)) \times p(m)$], a false rejectance rate of 0.67% was obtained. Sequential sampling with the optimum plan (2−30 ng/g for both 10 and 20 kg samples) predicted a false rejection rate of 0.21%, a >3-fold reduction. The false acceptance rate [obtained from computing $\sum_{m>15} p(\text{acceptance}|m) \times p(m)$] was reduced from 0.18 to 0.16%. Of course, all false negative rates are low because very few pistachio lots exhibit large aflatoxin test values.

Finally, one may compare the predicted rejection rates with those actually obtained by DFA for the 1998 crop: 1.4% of DFA samples tested >15 ng/g and would have been rejected on the basis of a single 10 kg sample. Had DFA used the proposed 2−30/2−30/15 optimal sequential test, 95% of samples would have led to acceptance, based on a test of ≤2 ng/g. Another 1.5% would have led to rejection based on a single sample >30 ng/g, whereas 3.5% would have required retesting with unknown results. A simple test of the sequential sampling proposed here would be to carry out the additional retesting of those 3.5% of pistachio samples at a laboratory such as DFA and comparing the sequential results with those reported based on single samples.

Comparison to Other Pistachio Distributions. Incidentally, the collection of 1293 10 kg tests of 1998 crop lots (Table 1) meets the requirements of a "sparse" test, as defined in Schatzki (1995a). Hence, it is possible

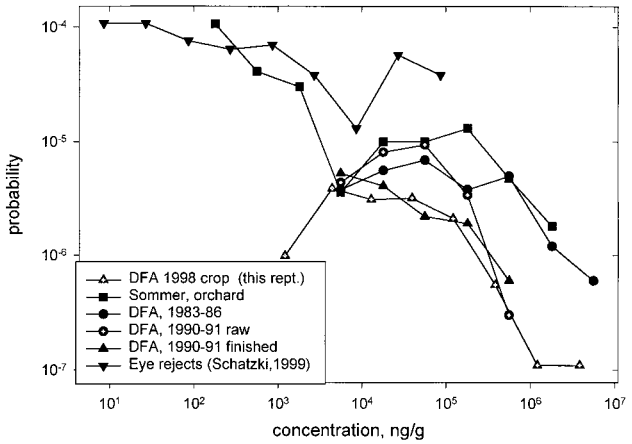

Figure 4. Pistachio single-nut aflatoxin distributions; probability of a nut falling in a half-decade concentration bin. Data are from Schatzki (1998) unless indicated otherwise.

to derive an individual nut distribution (p_i, c_i) among the 1998 tested lots. Following Schatzki (1995b), one bins the sample concentration given in Table 1 into logarithmic half-decade bins, designating the resulting probabilities as P_i and associating these with the bin midpoints C_i. Because this sample aflatoxin distribution corresponds to 10 kg = 7000 nut samples, one obtains the single-nut distribution p_i,c_i from $p_i = P_i/7000$ and $c_i = 7000 \times C_i$. The single-nut distribution may be compared with distributions derived previously from sample distributions of different sample sizes (Schatzki, 1998). This is presented in Figure 4. The agreement between the previously presented data and the present results is striking, particularly with the 1990 DFA finished nut data. Of course, the 1990 data again refer to an entire crop year and nuts ready for sale. This agreement reinforces the surmise, first stated in Schatzki (1995b), that these distributions are characteristic of pistachios as grown and processed and not of individual lots.

Conclusions. Use of three-step sequential sampling allows steeper OC curves in pistachio sampling for aflatoxin. With an acceptance level of 15 ng/g total aflatoxin, optimization of the regions of indifference to 2−30 ng/g for one and two 10 kg sample averages and a final decision at 15 ng/g for three 10 kg samples result in an OC curve approximating that of a 50 kg sample. This is a clear improvement over that of a single 30 kg sample obtained from a single test. Because the aflatoxin distribution in pistachios is very broad, the improvement to a virtual 50 kg sample is significant. Moreover, if the distribution of all lot aflatoxin averages is estimated from the test results of 1293 lots tested during the 1998 crop year, one predicts that fully 96.5% of all lots can be classified on the basis of a single 10 kg test. Use of the proposed sequential protocol effectively reduces the sample size to <10.7 kg (96.5% + 2 × 3.5% of 10 kg), requiring but 1.035−1.07 tests/lot.

ACKNOWLEDGMENT

I thank Dr. G. Ludwig of Schramm, William and Associates of Washington, DC, for suggesting the use of sequential sampling and Mr. Mike Hurley of DFA of California for making the 1998 pistachio test results available.

LITERATURE CITED

Schatzki, T. F. Distribution of aflatoxin in pistachios. 1. Lot distributions. *J. Agric. Food Chem.* **1995a**, *43*, 1561−1565.

Schatzki, T. F. Distribution of aflatoxin in pistachios. 2. Distributions in freshly harvested pistachios. *J. Agric. Food Chem.* **1995b**, *43*, 1566−1569.

Schatzki, T. F. Distribution of aflatoxin in pistachios. 5. Sampling and testing U.S. pistachios for aflatoxin. *J. Agric. Food Chem.* **1998**, *46*, 2−4.

Schatzki, T. F. Distribution of aflatoxin in pistachios. 6. Buyers and Sellers Risk. *J. Agric. Food Chem.* **1999**, *48*, 3771−3775.

Wald, A. *Sequential Analysis*; Wiley: London, U.K., 1947.

Whitaker, T. B.; Dickens, J. W. Simulation of aflatoxin testing plans for shelled peanuts in the United States and in the export market. *J. AOAC* **1989**, *72*, 644−648.

Received for review March 14, 2000. Accepted July 1, 2000.

JF000342O