Performance assessment under field conditions of a rapid immunological test for transgenic soybeans

John Fagan,* Bernd Schoel, Anne Haegert, John Moore & John Beeby

Genetic ID Incorporated, 1760 Observatory Drive, Fairfield, Iowa, USA

(Received 23 October 2000; Accepted in revised form 29 January 2001)

Summary Current market conditions and food regulations make it necessary for international and domestic participants in the agrifood industry to structure supply chains that control the content of genetically modified (GM) material in their products. Tests to detect and/or quantify GM components represent an important tool in maintaining such supply systems. This study assesses the field performance of kits that employ lateral flow immunotechnology to detect soybeans GM to be resistant to the herbicide glyphosate. Operators at 23 grain-handling facilities were paid to conduct analyses on a series of blinded samples containing defined proportions of conventional and transgenic soybeans. The observed rate of false positives was 6.7% in an experiment in which the highest level of GM material was 1% and 22.3% in a second experiment in which the highest level of GM material was 10%. This difference may be attributed to increased risk of cross-contamination with the higher level of transgenic material used in the second study. Samples containing 0.01% GM material were reported as genetically modified 6.70% of the time, while samples containing 0.1, 0.5 and 1% GM material were classified as genetically modified 29.5, 67.7 and 68.2% of the time, respectively. Thus, the frequencies of false negatives were 93.3, 70.5, 33.3 and 31.8% for samples containing 0.01, 0.1, 0.5 and 1.0% GM material. Samples containing 10% GM material were correctly reported as genetically modified in all cases. These results lead to the conclusion that the kit under study is useful in screening for lots of soybeans that contain high levels of GM material, but that, as a field tool, it is not effective in monitoring for GM material at the level of 1.0% or lower. Statistical and immunochemical analyses were carried out in order to assess the relative contributions of various factors to the error observed in these studies. These analyses indicated that limitations in operator performance, not defects in test kit materials, were the primary contributors, while sample size may play a secondary role. As both operator performance and sample size are independent of the specific characteristics of the test kit used in this study, it appears justifiable to generalize conclusions obtained here to other similar test systems.

Keywords Genetically modified organism, immunological analysis, lateral flow test, polymerase chain reaction, transgenic foods.

Introduction

Current market conditions and food regulations in many nations worldwide (for instance, Regulation (EC) No 258/97 of the European Parliament; Council Regulation (EC) No 1139/98, the Swiss Food Ordinance, SR817.0Z) make it necessary for international and domestic participants in the agrifood industry to provide information to clients regarding the presence of genetically modified organisms (GMOs) and their derivatives in food products. In most cases these requirements make it

*Correspondent: E-mail: jfagan@genetic-id.com

necessary to structure supply chains that control this characteristic and to conduct testing to quantify GMO content in the product (Hemmer & Pauli, 1998).

Two approaches are commonly employed in detection of GM materials in foods (Ehlers *et al*., 1997; Greiner *et al*., 1997; Hemmer, 1997; Hubner *et al*., 1999; Lipp *et al*., 1999; Meyer, 1999; Stave, 1999):

1 The polymerase chain reaction (PCR) can be used to detect and quantify transgenic genes or DNA sequences (McGarvey & Kaper, 1991; Meyer, 1995; Greiner *et al*., 1997; Hoef *et al*., 1998; Shirai *et al*., 1998; Wassenegger, 1998; Hubner *et al*., 1999; Lipp *et al*., 1999; Onishchenko *et al*., 1999; Vaitilingom *et al*., 1999; Vollenhofer *et al*., 1999).
2 Immunological methods can be used to detect transgenic proteins (Jankiewicz *et al*., 1999; Stave, 1999; Lipp *et al*., 2000).

Although direct measurement of transgene content offers advantages in terms of sensitivity, specificity, versatility and capability of providing meaningful quantification (Lindenmeyer *et al*., 2000), immunochemical methods, especially lateral flow or strip tests, would appear to offer advantages at certain early points in the production/supply chain, as they are fast, inexpensive and can be successfully completed in the field or in the grain-storage facility with minimal equipment. Immunological tests in the lateral flow configuration are offered by an increasing number of manufacturers and are currently available for detection of glyphosate-resistant soybeans and for those transgenic maize events that express *Bacillus thurengiensis* endotoxins Cry1A(b/c) and Cry9c. Manufacturers of these products provide data documenting the performance of these tests under specific conditions. These data imply that these tests should be highly effective in the field, capable of reproducibly detecting the presence of genetically modified (GM) material in a lot of soybeans or corn at levels as low as 0.3%. The research presented in this paper was designed to assess performance under actual field conditions, using one of the commercially available tests, the Strategic Diagnostics Inc. (Newark, DE, USA) lateral flow test for Roundup Ready© soybeans.

The conclusion that can be drawn from this study is that this test may serve a useful purpose early in the production chain as an initial rough screen for transgenics. However, because the sensitivity and reliability of this method are not sufficient to provide consistent discrimination at the 1% threshold, which is relevant to the EU and other regions of the world, this test must be used as part of a strong identity preservation system that incorporates more sensitive testing methods and additional quality assurance measures.

Materials and methods

Field test sites

Grain-handling facilities and grain analytical laboratories located in Iowa, Illinois, Indiana, Minnesota and Ohio were called randomly until twenty-four facilities were located that were willing to analyse a set of ten soybean samples. Inclusion in this study was based strictly on verbal assurance that the facility was routinely using the lateral flow tests that were the subject of this research project, and that the facility was prepared to conduct tests on soybean samples delivered to them. Twenty-two grain handling facilities and two state grain analytical laboratories were finally included in the study. Facilities were paid on a per-test basis at a rate that adequately compensated for time and materials.

Test materials

Several lots of soybeans were tested by PCR analysis to identify a lot that contained a very high percentage of transgenic soybeans. Two thousand seeds from this lot were planted and seedlings exposed to glyphosate (0.5%) at a stage at which the first foliage leaves had just opened. Of these, none were found to be sensitive to glyphosate. Using a binomial distribution model, we conclude that more than 99.5% of this soybean population is glyphosate resistant ($P \geq 0.999$).

Several lots of dark hilum conventional soybeans were subjected to PCR analysis to detect the presence of glyphosate-resistant transgenic beans. Samples of 10 000 beans were tested under conditions where a single transgenic soybean would have been detected. This level of sensitivity was verified using standards of known composition. One lot, shown in this initial screen to

contain no detectable GM material has been subjected to an additional nine tests, all of which were negative. Statistical analysis of these data indicate that the level of GM beans in this lot is, with $P \geq 0.999$, $< 0.01\%$.

Sample preparation

Appropriate amounts of the transgenic and conventional soybeans characterized above, were weighed out to yield 10 000 bean samples containing GM soybeans at the following levels: 0.0, 0.01, 0.1, 0.5 and 1.0%. In study no. 1, each grain handling facility was provided with a set of ten blind samples, two samples of each composition listed above. In study no. 2, each grain handling facility was provided with a set of four blind samples, one containing 0.0% GM soy and three containing 10% GM soy.

Sample submission and data collection

Sample sets were provided to each grain handling facility with instructions 'to carry out analysis according to the same procedure that is routinely used in your facility.' In study no. 1, the samples were provided in two five-sample lots to a few labs, but in a single ten-sample lot to most. In study no. 2, samples were provided to all nine grain handling facilities as a single lot. Results were reported in writing in the format routinely used at that facility. Samples containing GM soy were reported as '+', 'positive', 'GM' or 'detected'. Samples in which GM material was not detected were reported as '–', 'negative', 'GM free', or 'not detected.' Original reports are available for examination via the communicating author.

Of the twenty-four grain-handling facilities originally recruited for the study, twenty-one submitted full sets of data for study no. 1. One of the original facilities withdrew, one submitted data on five samples, and one submitted data on eight samples. The nine facilities recruited for study no. 2 all submitted full data sets.

Data analysis

All reports were converted to common terminology: '+', indicating GM material detected and '–', indicating GM material not detected. The data for study no. 1 is shown in Table 1 and summarized in Table 2, while the data for study no. 2 are presented in the text. Statistical analysis of these data was used to address the questions discussed in the following paragraphs:

Table 1 Results of study no. 1

Facility no.	Actual GMO content of samples (%)					Beans per sample
	0	0.01	0.10	0.50	1	
1	–	–	–	+	–	120
1	–	–	+	+	+	
2	–	–	–	+	+	70
3	–	–	–	+	+	60
3	–	–	–	–	–	
4	–	–	+	+	–	250
4	+	+	+	+	–	
5	–	–	–	+	+	100
5	–	–	–	+	–	
6	–	–	+	+	+	130
6	–	–	–	–	–	
7	–	–	–	+	–	90
7	–	–	–	+	+	
8	–	–	–	–	+	150
8	–	–	+	+	+	
9	–	–	–	+	+	100
9	–	–	–	–	–	
10	–	–	–	+	+	
10	–	–	–	–	–	
11	–	–	–	–	+	60
11	–	–		+		
12	–	–	–	–	+	180
12	–	–	–	–	–	
13	+	+	+	+	+	125
13	+	+	+	+	+	
14	–	–	+	+	+	2400
14	–	–	+	+	+	
15	–	–	–	–	–	50
15	–	–	–	–	–	
16	–	–	–	+	+	500
16	–	–	–	+	+	
17	–	–	–	–	+	250
17	–	–	+	–	+	
18	–	–	+	+	–	225
18	–	–	–	+	+	
19	–	–	–	+	+	100
19	–	–	–	+	–	
20	–	–	+	+	+	100
20	–	–	–	+	+	
21	–	–	–	–	+	250
21	–	–	–	+	+	
22	–	–	+	+	+	250
22	–	–	–	+	+	
23	–	–	–	–	+	
23	–	–	–	–	+	

Table 2 Summary of data reported in study 1

	Actual GMO content of samples (%)				
	0	**0.01**	**0.10**	**0.50**	**1**
Number of samples reported positive	3	3	13	30	30
Total number of samples	45	45	44	45	44
Percentage of GM samples reported correctly		6.70	29.50	66.70	68.20
Percentage of false negatives		93.30	70.50	33.30	31.82
Percentage non-GM samples reported correctly	93.30				
Percentage false positives	6.70				

First, a χ^2 test for the difference in binomial proportions (Collett, 1991) was conducted to determine if the proportion of false negatives reported in Table 2 differed significantly for samples with different levels of GM material (0.01, 0.10, 0.50 and 1%). The null hypothesis of the test is that the proportion of false negatives is equal for all samples. A significant result on this global test indicates that not all the proportions are equal but does not indicate which samples differ significantly. The latter question may be addressed by an 'analysis of means' (ANOM) for proportions (Ott *et al.*, 2000).

The analysis of means (ANOM) for proportions is analogous to a *post-hoc* multiple comparison in analysis of variance (ANOVA). The ANOM indicates which samples differ significantly from the overall pooled proportion of false negatives across the four samples. The null hypothesis for ANOM is that the proportion of false negatives for each sample is equal to the pooled proportion for all samples. Like multiple comparison procedures in ANOVA, the ANOM procedure controls the overall risk of type I error for the entire set of individual tests involved in the procedure.

In the two-tailed ANOM test procedure, if the observed proportion of false negatives is greater than the upper critical value for the test, we conclude that the sample in question is significantly larger than the pooled proportion of false negatives. If a sample proportion is less than the lower critical value it indicates that the observed proportion is significantly less than the pooled proportion of false negatives. The results of the ANOM are graphically displayed in Fig. 1. The upper critical value for the test is plotted as the upper decision line (UDL) in Fig. 1, the lower decision line (LDL) indicates the lower critical value, and the pooled average proportion for all samples is shown by the center line (CL) in Fig. 1. In the ANOM, the overall risk of type I error (α level) for the four individual tests taken together was set at 0.05.

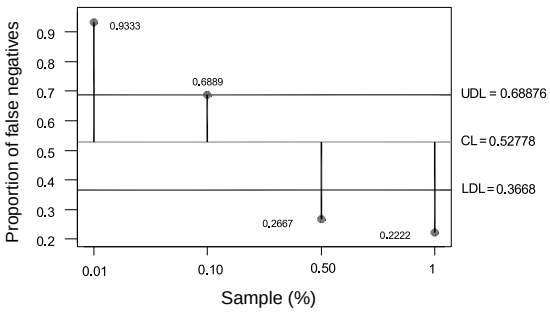

Figure 1 Analysis of means for proportion of false negatives. This figure presents results from the ANOM for the proportion of false negative tests for samples containing four different levels of GM material. The analysis indicates that the proportion of reported false negatives for the samples with 0.01 and 0.10% GM material were significantly greater than the overall pooled average proportion in all samples (shown by the centre line in the figure). These two samples correspond to the two points above the upper critical value (upper decision line or UDL). The two points below the lower critical value (LDL) indicate that the proportion of false negatives was significantly less than the pooled average for the other two sample categories. The overall α level for all four statistical tests taken together was 0.05.

In the ANOM for proportions, the upper and lower critical values (UDL and LDL) for the test were computed as follows (Ott *et al.*, 2000):

$$\mathrm{UDL} = \bar{p} + H_\alpha \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

$$\mathrm{LDL} = \bar{p} + H_\alpha \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

Table 3 Chi-square test for independence between results of tests 1 and 2 for samples with different levels of GMO content

Sample (% GMO)	χ^2 (1 d.f.)	*P*-value	Correlation coefficient (?)	Sample size
0.0	10.476	0.001	0.69	22
0.01	10.476	0.001	0.69	22
0.1	1.05	0.306	0.224	21
0.5	1.916	0.166	0.295	22
1.0	0.022	0.882	−0.032	21

where $\bar{p}$ is the pooled average proportion of false negatives across all samples, H_α is a tabled factor for a two-tailed ANOM with overall risk of type I error α and n is the sample size.

A second question addressed by the statistical analysis was the consistency or reliability of the results of the immunological tests administered by the grain-handling facilities. The results of this reliability analysis for study no. 1 are given in Table 3. The reliability of the test procedure was investigated separately for each different level of GM material (0.0, 0.01, 0.10, 0.50 or 1%). The statistical analysis employed data from all facilities that reported tests for each of the two blind samples with a given level of GM material.

Each of the two tests involved a binary outcome (+ or −), indicating the presence or absence of GM material. For each of the five levels of GM material, a χ^2 test of independence was used to determine whether there was a significant relationship between the binary outcomes of the two tests across all of the grain handling facilities (Bowerman & O'Connell, 1997). The null hypothesis of independence implies that the probability of a '+' (and '−') outcome should be the same for both tests.

Also the correlation between the outcomes of the two tests was calculated using the correlation coefficient ϕ, which measures the degree of association between two variables that are true dichotomies (Howell, 1997). The ϕ coefficient is the Pearson product–moment correlation for two dichotomous variables and it ranges in value from +1.0 to −1.0. Thus higher values of ϕ (in absolute value) indicate a stronger relationship between the outcomes of the two tests to detect the presence or absence of GM material.

For study no. 2, formal statistical tests were not necessary to analyse the reliability of the three tests administered by each grain-handling facility because all tests successfully detected the presence of GM material in the samples containing 10% GM material.

The third question addressed by the statistical analysis concerned the observed difference in the proportion of false positive test results for GM material reported in study no. 1 and study no. 2 (see Table 2 and text). Fisher's exact test for the difference between binomial proportions from two independent samples (Everitt, 1977) was used to determine whether the observed difference in the proportion of false positives was statistically significant. Fisher's test was used because diagnostic checks indicated that the size of the two samples was not large enough to support use of the normal approximation that underlies the usual test for the difference between two binomial proportions (Bowerman & O'Connell, 1997, p. 491–492).

Results

Study no. 1

Table 2 summarizes the results presented in Table 1, showing that samples containing 0.01% GM material were reported as genetically modified 6.7% of the time. Samples containing 0.1, 0.5 and 1% GM material were classified as genetically modified 29.5, 66.7 and 68.2% of the time, respectively. Thus, the frequencies of false negatives were 93.3, 70.5, 33.3 and 31.8% for samples containing 0.01, 0.1, 0.5 and 1% GM material. Samples containing 0% GM material were classified as containing GM material 6.7% of the time. Thus the frequency of false positives was 6.7%.

The data summarized in Table 2 were analysed statistically using a χ^2 test for the difference in binomial proportions. The latter test indicated that the overall difference in the proportion of false negatives across samples of differing GM content was statistically significant at the 0.05 level ($\chi^2 = 63.534$, 3 degrees of freedom, $P = 0.0000$, $n = 178$). This result indicates that the frequency with which grain-handling facilities failed to detect the presence of GM material differed significantly for the samples containing different levels of GM

material. To determine which individual samples were responsible for this significant overall results a two-tailed ANOM for proportions was conducted, with the overall risk of type I error controlled at 0.05.

The ANOM for proportions (Fig. 1) indicated that the proportion of false negatives for the samples containing 0.01 and 0.10% GM material were both significantly greater than the average pooled proportion of false negatives, although the proportion for the 0.10% sample was just barely significant. The proportion of false negatives was significantly less than the pooled average for the samples containing 0.50 and 1% GM material. As shown in Fig. 1, the UDL for the test was 0.6888, the LDL was 0.3668, and the pooled average proportion of false negatives was 0.5278. The test calculations were performed using MiniTab 13 for Windows. We conclude from this analysis that the frequency of false negatives was clearly related to the concentration of analyte in the sample at all analyte concentrations examined, from 0.01 to 1.0% GM material.

In the ANOM procedure, the average sample size was specified conservatively at 45. Thus the sample proportions used in the test (and reported in Fig. 2) were somewhat lower than those reported in Table 2 for the 0.10% sample (0.6889 vs. 0.7050) and 1% sample (0.2222 vs. 0.2270) which are based on the actual sample size in each case. Diagnostic checks indicated that use of the normal approximation to the binomial distribution required in the calculations of the critical values of the test statistic was justified for these data (Ott *et al.*, 2000, p. 327). The use of the normal approximation is appropriate in this case because both $n\bar{p}$ and $n(1-\bar{p})$ are >5.0.

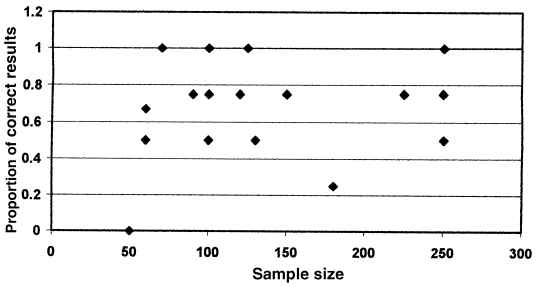

Figure 2 Lack of correlation between sample size and accuracy of analysis. The results in study no. 1 for analyses of samples containing 0.5% and 1.0% GM material were tabulated and the ratio of correct analyses to total analyses was calculated individually for each grain-handling facility. These values, which are a measure of accuracy of analysis, were plotted against the sample size used at that grain-handling facility. Two grain-handling facilities that used atypically large sample sizes – 500 and 2400 beans – were not included in the graph. Their accuracy scores were 1.0. Most grain-handling facilities analysed single sub-samples taken from each of the ten samples received for analysis (five samples for facility no. 2). However, a few facilities analysed two sub-samples from each of the samples received. For uniformity, these were tabulated for this graph as a single sample equal in size to the sum of the two sub-samples analysed. Only seventeen of the twenty three facilities provided adequate information on sample size to be included in this graph.

Assessment of test functionality

The high frequency of false negatives observed in Table 2 for samples containing 0.01 and 0.1% GM material would suggest that these levels are below the limit of detection of the test kit used, under field conditions. The frequency of false negatives observed with samples containing 0.5 and 1.0% GM material was still quite substantial. However, we suspected that these false negatives were more probably because of one or both of the following:

1 lack of consistent performance of test kit materials, or
2 specifics of the analytical procedure as performed by personnel at grain-handling facilities.

To assess the first possibility we conducted two laboratory assessments of lateral flow strips. In the first, a single sample of uniform analyte composition was prepared by extracting 20 g of powdered soybeans containing 10% GM material by weight, according to the kit manufacturer's instructions. This sample was distributed into 1.8 mL microfuge tubes in 0.5 mL aliquots. Into each tube a lateral flow strip was inserted and developed according to the manufacturer's instructions. A total of seventy-two strips were tested from three different lots. All tests yielded positive results, indicating 100% of the strips were functional with samples of this composition. Thirteen strips were also tested against an extract made from conventional soybean powder. All of

these analyses were negative. In the second assessment, forty strips were tested against a uniform sample containing 1% GM material. All tests yielded positive results.

These results indicate that the high frequency of false negatives observed in study no. 1 is not likely to be because of manufacturing defects in the lateral flow strip system. On the contrary, these results would suggest that the high level of false negatives is probably the result of specifics of the analytical procedure as performed in grain-handling facilities. These might include the following: (1) the sample size subjected to analysis might have been too small to provide statistically representative results; (2) the operator may not have conducted the tests correctly and/or consistently and (3) the operator may not have interpreted results of the tests correctly.

Study no. 2

To shed light on the relative importance of possible contributors to the observed performance of the lateral flow tests observed in Study no. 1, another set of samples was submitted for analysis to nine of the original grain-handling facilities. Each set consisted of one sample of conventional soybeans and three samples containing 10% transgenic soybeans. Each sample contained 10 000 beans. Operators were given the same instructions as for the previous study.

It was found that all grain-handling facilities detected the presence of GM material in all samples containing 10% GM soybeans, while the samples containing 0% GM material were reported positive with a frequency of two out of nine, or 22.2%. The frequency of false positives observed in this study was higher than that observed in Study no. 1 (22.2 vs. 6.7%). However, application of Fisher's exact test ($P = 0.190$) indicated that this difference, although practically or substantively significant, did not reach statistical significance, probably because of inadequate sample size. Power calculations revealed that, given the existing sample sizes, the power of the test was very low, with only a 21.4% chance of detecting a significant difference of 15.5% between the two proportions at the 0.05% significance level. This indicates that it would have been necessary to include a larger number of grain-handling facilities in the study for the difference in frequency of false positives between studies no. 1 and 2 to be statistically significant. It would not be surprising, however, to observe a higher frequency of false positives under conditions in which some samples in a series contain high levels of GM material (i.e. 10% or more). Under such conditions, cross-contamination during sample preparation would be more likely to introduce into a sample sufficient GM material to significantly increase its GMO content, thereby leading to a false positive result.

The results of this series of analyses indicate that grain-handling facilities can use this test kit effectively in identifying samples containing higher levels of GM material. Comparing the results of this study to those of study no. 1, where overall GMO content was lower, it was not possible to ascertain whether the higher level of false negatives observed in study no. 1 was a consequence of the sampling procedures used in the various grain-handling facilities or a consequence of other aspects of the analytical procedure.

To gain insight into this, we obtained information from participating grain-handling facilities regarding their sampling plan. The sample sizes used, which were consistent with the manufacturer's recommendations, are presented in Table 1. Figure 2 presents a scatter plot that examines the relationship between sample size and accuracy of analysis of samples containing 0.5 and 1.0% GM material. This figure clearly shows no correlation between these two variables. The lack of significant correlation would suggest a lack of strong linkage between the sampling plan used by these grain-handling facilities and the frequency of false negatives reported for samples containing 0.5 and 1% GM material. This suggests that, although sample size can certainly influence the frequency of false negatives, it is likely that other aspects of performance of the analytical procedure or data interpretation are important contributors, as well.

To assess this hypothesis more thoroughly, we statistically analysed the results of studies no. 1 and 2. In this analysis, we chose to examine the reliability of results obtained in repeated analysis to differentiate between sampling and procedural effects. The logic being that, if accuracy were limited by sample size, we would observe increased

reliability (as measured in terms of agreement between duplicate analyses) in analysis of samples containing increasingly higher levels of GM material, while if other factors were predominant, reliability would not necessarily be correlated with sample size.

Table 3 reports the results of the reliability analysis for study no. 1. The reliability of the test procedure (agreement between duplicate analyses) was investigated separately for each different level of GM content (0.0, 0.01, 0.10, 0.50 or 1%). The statistical analysis employed data from all facilities that reported results for two blind samples at a given level of GM content. The χ^2 test of independence employed in this reliability test requires that the expected frequencies for each cell of the 2×2 contingency table must be at least 5.0 (Bowerman & O'Connell, 1997, p. 1171), an assumption satisfied by these data.

As shown in Table 3, the χ^2 test indicated that the null hypothesis (i.e. independence of the outcome of duplicate immunological tests) could be rejected at the 0.05 level for samples containing 0.0 and 0.01% GM material. In contrast, the null hypothesis of independence could not be rejected for samples containing 0.1, 0.5 or 1.0% GM material. This means that operators at the participating grain-handling facilities failed to obtain consistent results when analysing duplicate samples containing GM material in the range from 0.1 to 1.0%. That is, there was no statistically significant correlation between the result obtained with one duplicate compared to the result obtained with the second duplicate when the actual levels of GM material in both samples was 0.1, 0.5 or 1.0%.

As sampling procedures were identical for replicates, the inconsistencies in results obtained for replicate analyses is not likely to be related to sampling limitations, but is more likely to be because of operator-related variability – variability in performance of the test or in interpretation of results.

Note that the results of the χ^2 tests for the 0.0 and 0.01% samples were identical. In each case, the correlation coefficient ϕ between the outcomes of duplicate analyses was 0.690. While the results of the statistical test were the same for both samples, the interpretation of the result differs for these two cases. In the case of samples containing 0.0% GM material, the correlation coefficient indicates that the grain-handling facilities were fairly consistent in correctly detecting the absence of GM material in those samples. In the case of samples containing 0.01% GM material, the correlation coefficient indicates that the grain-handling facilities were equally consistent in failing to detect the presence of GM material at the level of 0.01%. This result indicates that 0.01% GM material is clearly below the limit of detection of this test kit. In addition, the low correlations between the two test outcomes for the samples with 0.10, 0.50 and 1% GM content – with ϕ values of 0.224, 0.295, and -0.032, respectively – suggest that, as performed in grain-handling facilities, the test lacks adequate reliability for samples containing levels of GM material in this range.

Discussion

Testing for GM materials in foods can be used to meet three needs that arise along the production/shipping/storage chain. First, testing can be used at the primary receiving point in the chain, where production is pooled, to detect and eliminate truckloads that mistakenly contain substantial levels of GM material. Secondly, testing can be used at points in the chain where the product changes ownership to ascertain whether a given shipment contains a level of GM material that exceeds or falls below a certain threshold. This threshold is in some cases specified by law, as is the case for products destined for sale in Europe. In other cases, the threshold may be specified as one of the conditions of the contract between buyer and seller. The third need that testing can fulfill is to provide information as to whether a given product contains varieties of GM crops that have not been legally approved for use as food or feed.

Immunological tests, in both ELISA and lateral flow formats, have been developed and applied particularly to meet the first need mentioned above. Although studies have been published assessing the laboratory performance of ELISA tests (Jankiewicz *et al*., 1999; Lipp *et al*., 2000), the present study is the first to empirically assess the field performance of lateral flow tests.

The research presented here demonstrates clearly that the lateral flow test kit, studied here

under typical field conditions, is effective in detecting samples containing levels of GM material in the range of 10%, but is not effective in detecting lower levels (1% and below). Because operator performance, not the inherent characteristics of the kit materials, were found to be the primary factor influencing the field performance of this test, it is reasonable to expect that the results of this study can be generalized to other lateral flow kits designed to be used under similar field conditions to detect other targets relevant to commodities.

At the primary receiving point in the chain, the greatest risk to success in assembling a lot of product that contains little or no GM material is the inadvertent introduction of one or more truckloads of material with very high GMO content. For example, one truck containing 8 tonnes of 100% transgenic soybeans would bring the total content of GM material in an 800-tonnes silo to 1%. This would disqualify the lot for purchase by many buyers.

During harvest, truck drivers, farmers and operators of grain handling facilities may work continuously for several days with little rest and under substantial time pressure. Under such conditions the risk of mistakes, leading to accidental commingling, increases significantly. An on-site method for quickly checking the genetic status of each truck before it is unloaded serves as a highly useful cross check under such circumstances. In demonstrating that this kit can consistently detect GM material at the 10% level under field conditions, this study shows that lateral flow test kits can fulfil this purpose.

On the other hand, the results reported here indicate that use of lateral flow test kits carries substantial risk of false negative and false positive results if the grain-handling facility is operating to a 1% threshold for GM material. Under field conditions, 31.8% of samples containing 1% of GM material were mis-classified as not containing GM material. Thus, the classical 'buyer's risk' – false negatives – is substantial using this method. Similarly, these same conditions generated substantial 'seller's risk', as samples well under the 1% threshold, containing 0.5% GM material, were classified as GM 66.7% of the time. When operating to a 1% threshold, this constitutes a 66.7% frequency of false positives. Practically speaking, this means that under these conditions, product lots containing levels of GM material well under the 1% threshold will be rejected with frequency of 66.7%.

Both 'buyer's risk' and 'seller's risk' constitute serious hazards for the organization marketing grain or soy. If testing fails to detect the presence of 1% GM material in a shipment of product, the seller risks costly rejection of product when it is re-tested by the buyer. On the other hand, if testing classifies a sample as GM, when it contains only 0.5% GM material, a shipment of valuable product is unnecessarily rejected from the identity preservation system, leading to loss of value for the seller. We conclude that lateral flow tests alone, cannot be used with confidence to fulfil the second need posed above. If one wishes to consistently deliver products that comply with a threshold for GMO content of 1% or lower, lateral flow tests must be used in the context of a robust identity preservation system that incorporates the use of more sensitive and consistent analytical methods at critical points.

Compliance with thresholds for GMO content is required by a growing number of countries and must, therefore, be viewed as an important business objective at this time. For instance, EU labelling laws require that products containing more than 1% GM material be labelled as GM, while products with levels of GM material lower than this need not be labelled (Hemmer & Pauli, 1998). Non-compliance with this regulation constitutes grounds for removal of the product from the marketplace. Thus, for import into the EU, it is necessary to use an analytic method that can consistently distinguish samples that fall below this threshold from those that exceed it. Similar thresholds and import requirements are in force or are under development in many other countries, including Japan, Korea, Australia, Norway, Switzerland and colleagues.

Reliable answers to the third question posed above (whether a given product contains varieties of GM crops that have not been legally approved for use as food or feed) also require the use of more sensitive methods. This is because even the presence of 0.1% of an unapproved variety can be grounds for rejection of the product by the potential buyer, and because the presence of such levels of an unapproved GM variety in a food

product can constitute grounds for authorities to remove a product from grocery shelves in several countries. The results presented here indicate that lateral flow tests detect the presence of 0.1% GM material only 29.5% of the time. Thus, there is 70.5% likelihood that this method will fail to detect the presence of an unapproved crop variety, if it is present at the level of 0.1%. Clearly this method does not provide sufficiently stringent detection of unapproved varieties.

The results presented here suggest that two factors, sample size and operator performance, may limit the field performance of lateral flow tests for detecting the presence of GM material in food products. Two lines of evidence suggest that, although both of these may be important, operator performance appears to be the predominant determinant. First, as shown in Fig. 2, we did not find even a weak correlation between the stated sample size used by a grain-handling facility and the ability of that facility to provide accurate results. At each sample size, from fifty to 250 beans, accuracy scores varied widely, ranging from very low to the maximum possible score (1.0).

If sample size were the determining variable, we would have expected the range of accuracy scores to narrow as sample size increased. However, no such trend was apparent. Wide variability is not surprising in analyses that use very small sample sizes (50–60 beans), as the probabilities that a fifty bean sample will contain at least one GM bean are only 0.22 and 0.39, respectively, for samples taken from lots of beans containing 0.5 and 1.0% GM material. However, if sampling were a predominant determinant of accuracy under the conditions used by the grain-handling facilities, much stronger accuracy should have been observed with samples containing 250 beans than with samples containing fifty beans. This is because the probabilities that 250 bean samples taken from lots of beans containing 0.5 and 1.0% GM material will contain at least one GM bean, are much higher than the corresponding probabilities for samples of only fifty beans (0.71 and 0.92, respectively, compared with 0.22 and 0.39). Yet, we observed the same scatter of accuracy scores in analyses carried out with 250 bean samples as was observed in analyses made with 50 bean samples. This suggests that factors other than sample size are of primary importance in determining accuracy.

A second observation that also points to the importance of operator performance is presented in Table 3, which shows that no statistically significant correlation exists between results obtained with duplicate analyses for samples containing 0.1, 0.5 and 1.0% GM material. If sample size were a significant factor, we would have expected to see better agreement between duplicates for samples containing 1% GM material than samples containing 0.1% GM material. Data in Table 1 did not reveal such a correlation.

On the other hand, it should be pointed out that the results presented here do not eliminate sample size as a factor contributing to variability in results at grain-handling facilities. Although changes in sample size within the range used in most analyses reported in this study did not significantly affect accuracy, selecting very much larger sample sizes might improve accuracy. For instance, one facility included in the study used much larger sample sizes than other facilities (2400 beans). This facility achieved a perfect accuracy score. However, it would be necessary to analyse performance with many more samples of this size before we could confidently conclude that sample size was the determining factor, especially when it is noted that other facilities also achieved perfect accuracy scores with sample sizes of seventy five, 100 and 125 beans.

We conclude that, although some improvement in performance of the lateral flow method may be achieved by using larger sample sizes, operator performance appears to be a more important factor limiting the accuracy of this method under field conditions. This suggests practical strategies for achieving more reliable results for controlling the GMO content of foods. One approach would be to implement a more effective training programme for personnel who use lateral flow tests and to implement an effective quality control program designed to assure that tests are being performed to the desired standard. However, the nature of the field conditions and the variability in competency among those who conduct tests at grain handling facilities represent substantial challenges to implementing this strategy. An alternative approach would be to continue using the lateral flow test method as a screening method early in the food chain, with full recognition of its

limitations, and, in addition, to employ strong identity preservation procedures, including the use of more sensitive and consistent analytical methods, at critical points in the chain.

Acknowledgements

The authors thank Dr Kenneth Cavanaugh for statistical analysis and Dr Randy Coplin for his editorial contributions.

References

Bowerman, B. & O'Connell, R.T. (1997). *Applied Statistics: Improving Business Processes*. Chicago: Irwin.

Collett, D. (1991). *Modeling Binary Data*. Boca Raton: Chapman & Hall.

Council Regulation (EC) No 1139/98 of 26 May 1998 concerning the compulsory indication of the labeling of certain foodstuffs produced from genetically modified organisms of particulars other than those provided for in Directive 79/1, 12. /*EEC Official Journal L*, **159**(03106/1998), 4–7.

Ehlers, B., Strauch, E. & Goltz, M. (1997). *Fachbereich Genetik*. Nordufer 20, D-13353 Berlin, Germany: Gentechnik, Robert Koch- Institut.

Everitt, B.S. (1977). *The Analysis of Contingency Tables*. Boca Raton: Chapman & Hall.

Greiner, R., Konietzny, U. & Jany, K.D. (1997). Is there any possibility of detecting the use of genetic engineering in processed foods? *Zeitscrift für Ernahrungswiss*, **36**, 155–160.

Hemmer, W. (1997). *Food Derived from Genetically Modified Organisms and Detection Methods*. Clarastrosse 13 CH-9058 Basel, Switzerland: Agency for Biosafety Research and Assessment of Technological Impacts of the Swiss Priority Program Biotechnology of the Swiss Science Foundation. (BATS Report 2).

Hemmer, W. & Pauli, U. (1998). Labeling of food products derived from genetically engineered crops. *European Food Law Review of*, **1**, 27–38.

Hoef, A.M., Kok, E.J., Bouw, E., Kuiper, H.A. & Keijer, J. (1998). Development and application of a selective detection method for genetically modified soy and soy-derived products. *Food Addit Contam*, **15**, 767–774.

Howell, D.C. (1997). *Statistical Methods for Psychology*. 4th edn. Belmont, California: Duxbury Press.

Hubner, P., Studer, E. & Luthy, J. (1999). Quantitative competitive PCR for the detection of genetically modified organisms in food. *Food Control*, **10**, 353–358.

Jankiewicz, A., Broll, H. & Zagon, J. (1999). The official method for the detection of genetically modified soybeans (German Food Act LMBG 35). A semi-quantitative study of sensitivity limits with glyphosate-tolerant soybeans (Roundup Ready) and insect-resistant Bt maize (Maximizer). *European Food Research Technology*, **209**, 77–82.

'Lebensmittelverordnung' (Swiss Food Ordinance) of 1 March (1995). SR 817.02. Eidgenössische Drucksachen und Materialzentrale, CH-3003 Bern 1995.

Lindenmeyer, J., Hemmer, W., Auberson, L. *et al*. (2000). *Database on commercialized genetically modified foods, in Bgvv-Hefte 2000: Proceedings of the '4th DMIF-GEN Meeting 1999' at the Institute for Biological and Experimental Technology (IBET)*. Lisbon, Portugal, 5 October 1999. To be Published by the European Commission: Proceedings of the European Research Project SMT4-CT96-2072. http://food.ethz.ch/dmif-gen/lisbon

Lipp, M., Brodmann, P., Pietsch, K., Pauwels, J. & Anklam, E. (1999). IUPAC collaborative trial study of a method to detect genetically modified soy beans and maize in dried powder. *Journal of AOAC Internation*, **82**, 923–928.

Lipp, M., Anklam, E. & Stave, J.W. (2000). Validation of an immunoassay for detection and quantitation of a genetically modified soybean in food and food fractions using reference materials: Interlaboratory study. *Journal of AOAC Internation*, **83**, 99–927.

McGarvey, P. & Kaper, J.M. (1991). A simple and rapid method for screening transgenic plants using the PCR. *Biotechniques*, **11**, 428–432.

Meyer, R. (1995). Detection of genetically engineered plants by polymerase chain reaction (PCR) using the FLAVR SAVR tomato as an example. *Zeitscrift für Lebensm Unters Forschung*, **201**, 583–586.

Meyer, R. (1999). Development and application of DNA analytical methods for the detection of GMOs in food. *Food Control*, **10**, 391–399.

Onishchenko, G.G., Tutel'ian, V.A., Petukhov, A.I., Korolev, A.A., Aksiuk, I.N. & Sorokina, E. (1999). Current approaches to the evaluation of genetically modified food products. Soybean 40-3-2 data. *Vopr Pitan*, **68**, 3–8.

Ott, E.R., Schilling, E.G. & Neubauer, D.V. (2000). *Process Quality Control: Troubleshooting and Interpretation of Data,* 3rd edn. New York: McGraw-Hill.

Regulation (EC) No, 258/97 of the European Parliament and of the Council of, 27 January (1997) concerning Novel Foods and Novel Food Ingredients. *Official Journal of the European Communities*, **L 043** (14/02/1997), 1–7.

Shirai, N., Momma, K., Ozawa, S. *et al*. (1998). Safety assessment of genetically engineered food: detection and monitoring of glyphosate-tolerant soybeans. *Bioscience Biotechnology Biochemistry*, **62**, 1461–1464.

Stave, J. (1999). Detection of new or modified proteins in novel foods derived from GMO – future needs. *Food Contrology*, **10**, 367–374.

Vaitilingom, M., Pijnenburg, H., Gendre, F. & Brignon, P. (1999). Real-time quantitative PCR detection of genetically modified Maximizer maize and Roundup Ready soybean in some representative foods. *Journal of Agricultural and Food Chemistry*, **47**, 5261–5266.

Vollenhofer, S., Burg, K., Schmidt, J. & Kroath, H. (1999). Genetically modified organisms in food-screening and specific detection by polymerase chain reaction. *Journal of Agriculture Food Chemistry*, **47**, 5038–5043.

Wassenegger, M. (1998). Application of PCR to transgenic plants. *Methods Molecular Biology*, **92**, 153–164.