that a sample of 100 beans will contain a biotech bean and therefore a 37% probability that it will not. Therefore, if 10 sub-samples of 100 were taken from the same sample of 10,000 soybeans it is expected that three or four of the samples would *not* contain a single biotech bean and would result in a negative test, while six or seven of the samples would indeed contain biotech soybeans and would result in a positive response. The fact that both positive and negative responses are observed in the same sample is not an 'inconsistency' of the analytical method but the expected outcome and is certainly not because of 'operator-related variability'.

The authors further conclude that the test sensitivity is limited to concentrations above 1%. The test performed best at high concentrations in this study because most users have designed their individual sampling strategies for detection at these levels. Most of the users of these tests are testing soybeans at concentrations around the regulatory thresholds specified by their customers selling into Europe (1%) and Japan (5%) and therefore their sample sizes are designed around these screening levels – not 0.01 and 0.1%. Failure to use appropriate procedures to detect these low concentrations cannot be viewed as operator error because the labs were instructed by the organizers to use their 'normal sampling procedures'.

The sensitivity of the method is determined partly by the number of beans in the sample and the number of samples analysed from the load and can be adjusted to various levels of sensitivity with very high reliability. For example, if a person wanted to determine if a load contained 0.01% biotech soybeans, the analyst could test 10 samples of 1000 beans ground together, and providing the method is always positive when there is a single biotech bean in 1000, then 10 negative tests indicates that there was not a single biotech soybean in 10,000. This strategy can be employed to achieve any detection limit as long as the maximum number of beans in the sample is limited to a size where one biotech bean will always be detected. Ultimately, sensitivity of a method is usually limited by practical considerations like cost, time of analysis, etc. and not by the detection level of the analytical test.

The authors do point out that 'one facility included in the study used much larger sample sizes than the other facilities (2400 beans)'. This facility achieved a perfect accuracy score.

If the laboratories with the information that the authors intended to evaluate the test's capacity to detect at the level of 0.01, 0.1 or even 0.5% then the laboratories could have selected the correct sampling strategies to detect these levels. The authors explicit instructions to the laboratories and lack of information regarding the threshold screening concentrations, confidence levels and intended purpose of the study prevented the laboratories from using the test in a way that they could detect biotech beans in the blind samples.

In summary, it is our opinion that the experimental design of this study is flawed and as a result the experimental data generated does not support many of the conclusions as stated by the authors. This study does not add any substantial scientific information to the literature on biotech testing methods. It simply reinforces the necessity to choose an appropriate sampling strategy and testing method based on the particular testing application. No single sampling strategy or testing method can be used effectively for all applications but we do believe that immunoassay strip test method as specified by the USDA–GIPSA test method protocol is appropriate for the designated application.

G. David Grothaus PhD
President, AEIC Biotechnology Consortium
Leah Porter PhD
Executive Director of the Biotechnology Committee, American Crop Protection Association

Reference

Fagan, J., Schoel, B., haegert, A., Moore, J. & Beeby, J. (2000). Performance assessment under field conditions of a rapid immunological test for transgenic soybeans. *International Journal of Food Science and Technology*, **36**, 357–367.

Reply

Dear Editor,

I am writing in response to the comments made by Dr Grothaus of AEIC Biotechnology Consortium and Dr Porter of the American Crop Protection Association, on our article entitled, 'Performance assessment under field conditions of a rapid immunological test for transgenic soybeans', which was published in the April issue of this journal.

The respondents raised three major objections, which we discuss below:

Objection 1

The authors did not understand the methodology for the strip tests, and therefore the design of their study was not valid.

Quite to the contrary, it would appear that the respondents must have misunderstood the purpose and rationale of our study. It was not designed to evaluate the manufacturer's recommended protocol.

They have already completed that study. Instead the purpose was to evaluate how strip tests were being used under actual operating conditions in grain-handling facilities across the US. Such a field assessment of strip tests had not previously been carried out, and was needed.

In designing the study, we assumed that most grain-handling facilities operate to a pre-determined GMO threshold, typically 1%. This study was designed to assess how accurately strip tests perform in the region of that threshold under field conditions. We therefore provided a range of samples, from 0.01 to 10%, that bracketed the 1% threshold. These were presented to the facilities in a manner that modelled the kinds of samples that they usually receive, i.e., grain of unknown GMO content and unknown destination. This is why we used the term 'performance assessment under field conditions'. That elevators intend to operate to a routine threshold is a point upon which the respondents appears to concur. In his letter he states that operators use the manufacturers statistical model to select an appropriate sample size.

They add that 'most of the users of these tests are testing soybeans at concentrations around the regulatory thresholds specified by their customers selling into Europe (1%) and Japan (5%) and therefore their sample sizes are designed around these screening levels…We agree with the respondents' general assertion that grain-handling facilities routinely operate to a threshold that is defined by government regulations. However, at the time of our study (Autumn, 1999), the Japanese government had not set a threshold. It was not announced until December (1999). Thus, at the time of this study, the only established regulatory threshold was the 1% EU standard, and it was to this threshold that most users were operating. This was borne out by our interactions with grain handling facilities at that time.

Objection 2

The respondents contend that the grain handling facilities used inadequate sample sizes because the authors did not provide them with sufficient information regarding the levels of GM material in the samples.

The respondents state, 'The authors did not tell the labs what concentrations to screen for, but instead instructed the labs to use their 'normal sampling procedure'. It is our understanding, in practice, that each user of the test screens at different concentrations determined by their unique business considerations… The respondents imply that grain-handling facilities generally customize their sampling and testing methodology on a per-lot basis, depending on destination requirements. They conclude that, in order to assess a facility's routine performance, a threshold and confidence interval must be specified, and they state that we failed to provide that information. The following four points establish that grain-handling facilities generally operate to a routine analytical procedure.

1. The very nature of grain-handling operations precludes the respondents' contention that facilities adjust test methods on a per-lot basis. At the time of receipt, when a particular lot of grain is tested for GMOs, the facility frequently does not know who will be the final customer for that specific lot or where it will be shipped. Therefore, they cannot know what threshold or confidence level would be specifically relevant to that lot. Thus, it is often impossible for them to customize their testing methodology on a per-lot basis. Instead, facilities establish a routine sampling and testing protocol that is applied uniformly.
2. The respondents' contention that facilities alter test methods on a per-lot basis directly conflicts with the manufacturer's statement, elsewhere in his letter, that facilities usually test lots to a predetermined regulatory threshold (see response to first objection, above).
3. When we asked facilities to test samples for us, we did not specify testing parameters. Rather, we simply asked if they could test to determine whether our soybeans were non-GMO. Not a single facility asked us the questions that, according to the respondents, were 'an absolute requirement of the methodology.' If the facilities had been routinely operating to a variety of thresholds they certainly would have asked us what threshold and confidence interval would be required for our purposes. On the contrary, it was clear from our discussions with them that these facilities had established routine sampling procedures that they used to determine whether a product was non-GMO.
4. Similarly, in reporting strip test results, not a single facility specified a threshold or a confidence interval. They simply reported GMO positive or negative. If the facilities routinely customized their tests, they would have specified thresholds and confidence intervals in their reports.

If a grain-handling facility routinely operates to a certain standard (e.g. 1%), then that facility has already established a routine threshold and confidence interval. Requesting that they use their standard procedure, as we did in organizing this study, implicitly specifies the threshold and confidence interval that pertain to that standard procedure. Thus, the respondents were not correct when they claimed that we failed to provide the grain handling

facilities with sufficient information to conduct valid testing. We implicitly provided that information when we requested that they operate to their normal or standard procedure.

The performance of that standard protocol defines what the grain handling facilities mean by 'GMO positive' and 'GMO negative.' The accuracy and consistency of those operational definitions are what was evaluated by our study.

Objection 3

The respondents contend that the high rate of false negatives that was observed was strictly because of the use of inadequate sample sizes.

This was, in fact, an early hypothesis in our research. However, it was rejected based on the analysis presented in the study, some of which is discussed below.

1. In their argument claiming that the high rate of false positives was because of inadequate sample size, the respondents focus exclusively on a sample size of 50 beans. However, only one facility out of 23 used a sample size of 50. Sample sizes ranged from 50 to 2400 beans, with only five of the 23 facilities using sample sizes of less than 100 beans. Thus, the respondents' analysis does not adequately address the data presented and leads to false conclusions.

   [It is interesting to note that the respondents attempted to defend the facility's use of such a small sample size by claiming that they must have been working to the Japanese 5% threshold. That threshold, however, had not yet been announced by the Japanese government.]

   If one examines results obtained with the sample sizes representative of the majority of grain handling facilities in the study, one readily comes to the conclusion that sample size may contribute to, but does not account fully for the high frequency of false negative results observed in the study.

   For samples containing 225–250 and 90–130 beans, the expected rate of false negatives, based on sampling statistics alone, are 9.1 and 29.5%, respectively. In contrast, the observed frequency of false negatives was 20% for samples of 225–250 beans, which is over twice the frequency expected based on sampling statistics, and 37.5% for samples of 90–130 beans, which is also significantly greater than the expected frequency. Certainly the reduced frequency of false negatives in the larger sample size suggests that sample size does contribute somewhat to reducing false negatives. However, the large differences between observed error rates and error rates expected based on sample size calculations indicates that factors in addition to sample size had a significant influence.
2. One of the key observations that implicated factors other than sample size was the lack of correlation between sample size and the frequency of agreement between analyses of duplicate samples, as sample size increased from 50 to 250 beans. the respondents challenged this by stating that, on the basis of statistical calculations one would expect that 37% of samples of 100 beans, taken from a lot containing 1% GM soy, would not contain a single GM bean, and would therefore come up negative. While this is certainly the case, the critical point that he seems to have missed is that our analysis was not based on examining duplicates of a single sample size. Instead, it compared the extent of agreement between duplicates of different sample sizes. Statistical analysis would suggest that, as sample size increases from 50 to 100–250 beans, the percent of replicate samples that contain at least one GM bean should increase progressively; whereas 63% of 100 bean samples should contain at least one GM bean, 92% of 250 bean samples should contain at least one GM bean. Thus, based on statistical analysis, as sample size increases, the frequency with which duplicate analyses agree should increase. The empirical results did not agree with these statistics-based expectations. Little correlation was observed between sample size and agreement between duplicates, indicating that factors other than sample size contributed to the high frequency of false negatives.

In their final remarks, the respondents refer to the 'USDA-GIPSA test method protocol'. This is puzzling, since (1) this protocol was not in force at the time our study was carried out, and (2) this protocol is not relevant to the strip test that was the subject of our study (RoundUp-Ready specific), but applies to a strip test for another GMO, StarLink© maize. However, it is worthwhile to mention that we have recently carried out a field performance evaluation of StarLink strip tests, and find similarly high rates of false negatives, rates 5-fold greater than the rates attributable to sampling statistics, alone. Thus, results with this second strip test confirm those reported in our published study.

In closing, the respondents seem to have misunderstood the purpose of our study. The results and conclusions of our paper do not challenge the quality of the immunoassay kit or the validity of its prescribed sampling methodology. Rather, the study reveals that, in the field, as used in grain-handling facilities, the performance of these strip tests is much

less accurate than would be expected based on their laboratory performance and on statistical sampling models.

In the study, we evaluated the factors that might contribute to this reduced quality of performance. We ruled out defects in the test technology and sample size as primary factors, concluding that the remaining variable, operator performance, was key. Because strip tests accurately detected 10% GM content in the field, the study also concluded that strip tests are a useful tool for screening for high concentrations of GMOs, and that when operating to lower GM thresholds, such as 1%, strip tests should be used in the context of a robust identity preservation system, which employs other, more sensitive and accurate testing methods at critical points.

Sincerely,

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

Reference

Fagan, J., Schoel, B., Haegert, A., Moore, J. & Beeby, J. (2000). Performance assessment under field conditions of a rapid immunological test for transgenic soybeans. *International Journal of Food Science and Technology*, **36**, 357–367.

Letter to the editor

Dear Editor,

The study reporting the alleged inaccuracy of genetically modified (GM) strip testing on soybeans (Volume 36, Issue 4, April 2001) unnecessarily raises alarming questions about the integrity of American agriculture and its ability to reliably market 'non-GM' grain reliably.

The article, written by Genetic ID employees of Fairfield, Iowa, cited a high incidence of false results for GM-detecting tests incorporating lateral flow immunotechnology. While the credibility of the study has been questioned by test strip manufacturer Strategic Diagnostics, Inc., it is more important to note the limitations of the study's scope, which focuses on only one methodology (i.e., testing) for determining the presence of GM material in US grain. There is a more comprehensive solution, briefly alluded to at the end of the article as 'strong identity preservation procedures.' It is called traceability, a term denoting a field-to-food identity-tracking system.

Traceability systems are designed to preserve the integrity of specialty and segregated agricultural products. Right now, it is possible to know the exact origin of the grain purchased by merchandisers or processors, just by reviewing data collected at critical points in the food chain, via ISO 9000 systems. Through crop auditing companies like ours (CropVerifeye.com, LLC), grain buyers can review audited data from seed inspections, variety verification, equipment and bin sanitation, isolated field row documentation and field monitoring for weeds, disease and insect damage and off-type plants. In other words, grain buyers around the world have the power to know-through the use of CropVerifeye.com-online, in real time, whether the grain they are considering for purchase is originally GM or non-GM or whether its identity has been preserved as such. This validation of identity preserved grain can then extend on through shipping, delivery and processing, if the grain buyer requests electronic traceability beyond the field.

We need to shift our attention from isolated incidences of alleged GM testing inaccuracies and concentrate on the bigger picture. Do we really want to entrust the United States' presence in the global marketplace to the performance of a few GM testing tools?

We must know what kind of grain we have, from the day the seed is planted to the day the finished product finally rolls off the conveyer belt at the processing plant. Within this context, GM tests assume their correct role, which is ancillary.

To truly safeguard the future of US exports and the satisfaction of consumers around the world, we need a seamless audited tracing system, from field to food.

Jim Mock PhD
CropVerifeye.com, LLC, Wichita, Kansas, USA