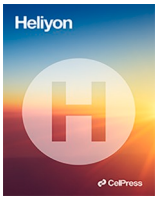
Heliyon 10 (2024) e34768

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Field validation as a means for continual monitoring of approved test kit's fitness for purpose in the commercial market

Megan K. Rooney *, Timothy J. Herrman

Office of the Texas State Chemist, Texas A&M AgriLife Research, College Station, TX, 77841, USA

ARTICLE INFO

Keywords:
Fumonisin
Test kits
Toxicology
Mycotoxins
Risk management
Regulation
Statistics
Maize
Validation
Quality systems

ABSTRACT

Testing accuracy of a chemical contaminant requires use of a testing platform that conforms to validation criteria outlined in quality literature and standards. This study explores the application of commercial field data measured by qualified analysts using a United States Department of Agriculture – Federal Grain Inspection Service approved kit for measuring fumonisin in maize to augment method validation procedures. Analysts from seven grain testing facilities were qualified in official USDA sampling, sample preparation, and testing methodology using the Charm LF-FUMQ-WETS5. A duplicate sample was tested in the Office of the Texas State Chemist (OTSC) laboratory using UPLC-MS-MS. Data were subject to four statistical techniques using continuous and categorical methodology. This approach enabled researchers to explore if a single test or multiple comparisons were best suited to assess a field kit's fitness for purpose across facility, toxin level, and year. The study concluded that a paired *t*-test and correlation analysis provided a quick and meaningful evaluation of kit performance. The correct placement of samples within the correct bin (violative versus non-violative) aligns well with market forces and regulatory compliance. The results of this study also provide a useful tool to assess all field kits' performance at the beginning of the harvest season and subsequent years. The combination of statistical techniques presented in this research is an important tool in assessing mycotoxin field test kits fitness for purpose and represents a key step in a continuous improvement-quality systems process meant to protect the feed and food supply.

1. Introduction

Validation of laboratory measurement procedures including sampling, sample preparation, and testing methods is a criteria within the ISO/IEC 17025 standard and a core function of some associations including AOAC International. The Office of the Texas State Chemist (OTSC) is currently accredited under ISO/IEC 17025:2017 ("General requirements for the competence of testing and calibration laboratories") for chemical and biological methods, and their scope of accreditation include Federal Grain Inspection Service (FGIS) performance verified test kits for analyzing maize samples for aflatoxin. OTSC collaborates closely with the FGIS technical laboratory in communicating the testing needs of the Texas Feed and Fertilizer Control Service (here after referred to as the Service). Included in these collaborations are the expansion of the testing range for aflatoxin [1] and fumonisin test kits. The FGIS performance testing criteria requires firms to submit a complete data set specified in the guidance document [2], and the FGIS technical staff duplicates this analysis prior to approval. These lists of approved kits are updated on a regular basis to include the addition of new

* Corresponding author.
E-mail address: mkr@otsc.tamu.edu (M.K. Rooney).

https://doi.org/10.1016/j.heliyon.2024.e34768
Received 28 November 2023; Received in revised form 15 July 2024; Accepted 16 July 2024
Available online 17 July 2024

technology and to remove kits that no longer meet performance criteria. The FGIS advisory committee provides input on the testing range they deem important for their industry. The Service also provides input on the needs of the Texas grain industry.

Prior to 2011, the testing range for FGIS approved aflatoxin kits was 100 μg/kg (ppb). It is not uncommon for Texas to experience aflatoxin levels exceeding 300 μg/kg, and by rule, the Service approves use of maize containing up to 500 μg/kg for blending with maize containing greater than 20 μg/kg for use in feed not to exceed 200 μg/kg [3]. Following aflatoxin test kit validation to 500 μg/kg by OTSC, FGIS subsequently expanded their validation of these kits to 300 μg/kg [4]. Prior to 2016, the FGIS testing limit for fumonisin test kits was 5 mg/kg (ppm). In 2016, FGIS expanded the testing criteria to 30 mg/kg. This upper range was established in consultation with the FGIS advisory committee. The Service provided separate input to the FGIS technical office indicating the need for kits to perform over 60 mg/kg based on the historically high levels of fumonisin that occur in the Texas High Plains. The Service had performed an independent validation for the one field kit that claimed to test up to 200 mg/kg using reference material with a 74 mg/kg fumonisin contamination and had approved this kit for testing up to 100 mg/kg for use in Texas [5]. FGIS subsequently approved the use of this kit with an upper testing range of 60 mg/kg for use in Texas by their approved labs in the Texas High Plains in 2017. While this research focuses on field validation of the only approved fumonisin kit available at the time of this study, the techniques have broad implications about how the Service constantly monitors field kit testing performance. For example, on several occasions in Texas, FGIS approved kits were removed from the Texas market for failing to perform under their FGIS design criteria and test performance specifications. In these instances, the Service communicates their finding to the test kit manufacturer to help resolve the issue and have the firm voluntarily remove their product from the TX market. As an outcome of these situations, there is a need for the Service to establish field validation criteria that they can follow as well as a means to inform the industry, FGIS, and other state regulatory agencies when a test kit is no longer performing according to specification.

The regulation of mycotoxins involves multiple agencies within the United States (US). For example, the US Food and Drug Administration (FDA) first established aflatoxin action levels in 1969 and has subsequently revised these as new science supports their revision [6]. Similarly, the FDA established fumonisin guidance levels in 2000, which were updated in 2001 [7]. While the USDA has no direct regulatory authority over aflatoxin, as a marketing agency, they are tasked with grain grading through the FGIS and perform mycotoxin testing upon request as an informational factor. They also test all export loads of maize shipped by vessel for aflatoxin unless the buyer specifically waives this requirement. Several states inculcate regulatory limits for mycotoxins in agricultural goods through promulgating rules, as has occurred in Texas for aflatoxin. Texas fumonisin rule TAC §61.61(a)(7) [8] pre-dated FDA guidance levels by a year but were subsequently removed from their rule in 2018 in preference for FDA guidance levels at the request of stakeholders (OTSC Advisory Meeting, October 5, 2018).

Fumonisin is a mycotoxin produced by several *fusarium* mold species. Fumonisin is toxic to equine, causing leukoencephalomalacia, and to swine, causing pulmonary edema [9]. In humans, fumonisins are linked to esophageal cancer and neural tube defects through epidemiological studies [10–12]. US FDA guidance levels for humans is 2 mg/kg and 5 mg/kg for animals with an additional stipulation that it not exceed 50 % of the ration, except in horses and rabbits where it may not exceed 20 % of the total ration [7]. The FDA guidance provides higher levels for some animal classes and growth stages. For example, maize containing up to 60 mg/kg (ppm) can be fed to cattle past the immature stages and to poultry up to 100 mg/kg (ppm) during grow out with an inclusion limit of 50 % of the ration. The significance of these levels with regard to testing and purchasing maize is that the FDA upper limits for these uses exceed the FGIS verified performance criteria. There exists a need for expanding kits' testing capability, particularly for the Texas market. As previously discussed, the Service and FGIS worked collaboratively to address this issue.

An essential component to test kit validation is the use of control samples. ISO published a standard for the creation of reference material (ISO 17034:2016). Texas is an ideal location to source highly contaminated cereals and oilseeds for the production of reference material for aflatoxin and fumonisin contaminated maize. OTSC received ISO 17034 (2016) accreditation for the production of both aflatoxin and fumonisin ground maize reference materials [13]. The use of this reference material is an important addition to monitoring testing accuracy by Texas firms and facilitating test kit validation. Previously, OTSC utilized similar protocol to the reference material production without the ISO accreditation [1].

A need exists to quantify performance criteria for field validation of mycotoxin test kits in Texas. This study explores a means for the continual monitoring of test kits' fitness for purpose in the commercial market. Therefore, the purpose of this research was to explore different statistical techniques including regression, matched pairs analyses (paired *t*-test), and categorical analyses and to formalize a field validation procedure that can be utilized by the Service.

2. Materials and methods

2.1. Analysis of fumonisin by UPLC-MS-MS

For all fumonisin determination by the lab, an ISO/IEC 17025 accredited in-house method was used for analyses. This method is based on the method developed at OTSC [14]. Modifications to the UPLC-MS-MS procedure included the following: extraction of a 50 g sample with 250 mL of 70:30 methanol:water; shaker time of 60 min instead of 15 min, and filtration of extract through Whatman #1 filter instead of centrifugation. Fumonisin certified analytical standards and isotope-labeled standard were purchased from Biopure-Romer Lab, Inc., Tullin, Australia (FB_1, 50 μg/mL, Cat. No. 002003; FB_2, 50 μg/mL, Cat. No. 002004; and FB_3, 50 μg/mL, Cat. No. S02007; fumonisin internal standard (U-[$^{13}C_{34}$]-FB_1), 25 μg/mL, Cat. No. ILM003). All commercial standards were solutions prepared in acetonitrile-water (50:50).

The validation of the in-house method for fumonisin determination included analysis of blank samples (cornmeal as matrix) fortified at different levels, regulatory samples, and working control samples previously tested by another in-house method (Protocol

16701 -"Determination of Fumonisin in Feed by HPLC using NDA"). Accuracy and precision target limits were taken from AOAC (Horwitz) [15,16]. Precision or repeatability is calculated as the relative standard deviation (RSD) (coefficient of variability), and accuracy is calculated as % recovery.

2.2. Reference materials

Naturally, contaminated corn was ground through a 0.75 μm screen on a Retsch SR300 and blended for 120 min in a commercial mixer (Multiquip, Model MC94PE) to produce the reference material with the desired fumonisin levels. These levels were confirmed using UPLC-MS-MS as a standard testing approach. Homogeneity and stability of material were established by measuring 12 samples in duplicate. Statistical design was taken from ISO 13528 for homogeneity testing [17]. The acceptable testing variances were set forth in the USDA-AMS-FGIS mycotoxin test kit specifications [2].

2.3. Kit validation

Charm Sciences, Inc. ROSA® WETS5 fumonisin quantitative test was validated by following the FGIS protocol [18]. Because the test kit had been validated at levels of 0.5 mg/kg, 2 mg/kg, 5 mg/kg, and 30 mg/kg by FGIS, the Service only validated at a level >60 mg/kg using 74 mg/kg reference material. A supplemental protocol, provided by Charm Sciences [19], was necessary for measuring above 40 mg/kg and included an additional dilution step. Three OTSC analysts conducted the test seven times using reference material. Data were analyzed using Excel to establish an extended performance range.

Firms analyzed samples using the Charm Sciences, Inc. ROSA® WETS5 Fumonisin Quantitative Test by following the instructions provided by FGIS [18] and the supplemental protocol (for results >40 mg/kg) [19]. Analysts were trained how to use the kit, and they were qualified by analyzing reference material samples provided by the Service. The requirements for qualification included performing two analyses with results duplicating within acceptable limits [20]. Criteria for developing the firm's annual sampling and testing plan are contained in the One Sample Strategy (OSS) Handbook [20].

2.4. Data analysis acceptance criteria

Data were analyzed using regression (Criterion 1), paired *t*-test (Criterion 2), FGIS range compliance (Criterion 3), and categorization of violative/non-violative results between OTSC and the facilities (Criterion 4).

Criterion 1. Regression analysis was completed on log transformed data. A Pearson's correlation coefficient (r) was used to explain the strength and direction of the relationship and amount of variance (r^2) between the firm and Service results.

Criterion 2. A paired *t*-test was used to compare the kit with the validated UPLC-MS-MS method at a 95 % confidence level. A test kit would "pass" if the p-value was greater than 0.05 (meaning no significant difference between the test kit and standard instrument method). Data were sorted into bins of <30 ppm, ≥30 ppm ≤ 60 ppm, and >60 ppm to further assess kit performance at various concentrations (ppm = mg/kg).

Criterion 3. A qualitative categorization comparison was performed to assess whether kit results placed the test sample in or out of FGIS compliance versus the "official result." If the test kit result fell within the FGIS acceptable range, then test kit result "passes". For example, if the "official" result (i.e. OTSC instrument result) was 10 ppm, then the test kit result would need to fall in the acceptable range of 7.4–12.6 ppm (Table 1). An overall 90 % pass rate (determined by [number of samples within range/total sample number] x 100) was set as the requirement for validation of the test kit. An additional assessment was conducted based on concentration categorization. Data were sorted into bins of <30 ppm, ≥30 ppm ≤ 60 ppm, and >60 ppm to assess kit performance at various concentrations (ppm = mg/kg).

Criterion 4: Categorization of violative/non-violative results between OTSC and the facilities provided information on Type I and Type II error and percent agreement (combined agreement on classification of non-violative and violative). The test kit would "pass" with 90 % agreement and ≤1 % Type II error.

JMP®, Version 17.1. AS Institute Inc., Cary, NC, 1989–2023 and Microsoft Excel (2016) were used for statistical analyses.

Table 1
FGIS acceptable performance criteria.

Fumonisins[a] (ppm)	Maximum RSD (%)	Standard Deviation (ppm)	Acceptable Range (ppm)[b]
0.5	18	0.090	0.32–0.68
2.0	14	0.28	1.4–2.6
5.0	13	0.65	3.7–6.3
30	13	3.9	22–38

For concentrations 5.0 ppm and above, the maximum RSD = 13 %.

[a] ± 15 %.

[b] $Range = Concentration \pm [(2 \times RSD) \times Concentration]$.

Source [2].

2.5. Data sets

The development of a field validation process came from the assessment of data from 2016 to 2017 crop year corn at six and seven establishments, respectively. Quality measures established for this study were provided by the One Sample Strategy (OSS) program [20]. Specifically, analysts were trained and subsequently qualified in sample collection, preparation and testing. During harvest, analysts ran reference material weekly, or more often if there was a change in the test kit lot. These data were recorded as well as incoming trucks, and reported to the Service weekly. These results were reviewed by Service personnel. Onsite inspections were performed at a minimum of three times and included collection of retained material (verification samples) as a further mechanism of validating analyst accuracy. These collected verification samples allowed for testing of the same ground material by the OTSC lab. The difference between the 2016 and 2017 harvests involved collecting outbound blending samples in 2016 and incoming loads in 2017 as well as outbound blended loads.

3. Results and discussion

The results of fumonisin test kit using the 74 mg/kg reference material are presented in Table 2. These results are within the theoretical prescribed range of testing using FGIS criteria for levels of fumonisin at 30 mg/kg and above (Table 1). These test kit validation results supported the decision by the Service to expand the acceptable testing limit of the Charm fumonisin test kit to 100 mg/kg.

For the UPLC-MS-MS validation results, the intra-day and inter-day reproducibility were evaluated at 200, 500 and 1000 μg/kg (ppb) for FB_1 and FB_2, and 100, 250 and 500 μg/kg for FB_3. In animal feed cornmeal, matrix recoveries of FB_1 ranged from 93 % to 98 % (RSD from 5 % to 8 %). Recoveries of FB_2 ranged from 104 % to 107 % (RSD from 2 % to 6 %). Recoveries of FB_3 ranged from 94 to 96 % (RSD from 2 % to 5 %).

Regulatory samples had concentrations of total fumonisin of 2, 13, and 64 mg/kg (ppm). Working controls had a concentration of around 6 mg/kg (ppm). Recoveries of FB_1 ranged from 95 % to 114 % (RSD from 2 % to 8 %). Recoveries of FB_2 ranged from 102 % to 107 % (RSD from 3 % to 7 %). Recoveries of FB_3 ranged from 93 to 106 % (RSD from 5 % to 8 %). By the application of an isotope labeled internal standard, the matrix effect in UPLC-MS-MS analysis was effectively eliminated, and the performance of quantification method met the requirements outlined in AOAC methodology as well as FDA and EU regulation criteria [21]. All correlation coefficients for the standard curves were greater than 0.99.

3.1. Quality Measures/Record Review

2016 Crop (blend plans). All six elevators were able to achieve the duplication limit for the provided reference material (Appendix 1). There was a variation in the number of times the reference material was analyzed by a facility. Overall, the material was run a total of 25 times with 100 % compliance to the duplication limit (acceptable range) for the 47 ppm control. A 30 % duplication limit was given due to the requirement for a fourth dilution (supplemental dilution) for quantitation above 40 ppm for this particular test kit. Particle size of 70 % fines through a 20 mesh sieve was met by all facilities (requirement set by the OSS Program [20]).

2017 Crop (OSS facilities). All seven facilities were able to achieve the duplication limit for the provided reference materials and particle size requirement (Appendix 2). Because these facilities were part of the OSS program, there was an additional stipulation that control material be tested a minimum of 20 times. Three levels of reference material were available: 42 ppm, 47 ppm, and 63 ppm. All facilities met the criteria of 20 controls run, and with the adjustment for the 63 ppm to a 30 % duplication limit (originally was set at 20 % for OSS program), all met the 90 % within duplication limits. Overall, controls were run 259 times with a compliance of 93.4 %. With the duplication limit adjusted for the higher control, the percentage went to nearly 100 %. All facilities met the requirement for particle size for the grinders. The requirement was for greater than or equal to 70 % of the ground particles (% fines) pass through a 20 mesh sieve.

3.2. Data analysis of verification samples: comparison of field results with OTSC results

The retained samples, referred to as verification samples and collected by the Service's investigators, were analyzed by the OTSC

Table 2
Validation results and quantitation range for Charm WET-S5 Kit using OTSC reference material (O2017-000072).

	Charm WET S5 Kit	UPLC-MS-MS Analysis
Mean	77 ppm	74 ppm
SD	9 ppm	5 ppm
RSD	12 %	7 %
Accuracy	4 %	
$\%Relative\ Standard\ Deviation\ (RSD) = \frac{Standard\ Deviation\ (SD)}{Mean} x\ 100$		
$\%Accuracy = \frac{(Mean\ of\ Reference\ Material - Mean\ of\ Test\ Kit)}{Mean\ of\ Reference\ Material} x\ 100$		

lab using the ISO/IEC 17025 accredited in-house method for fumonisin. The verification samples were the same material tested by the facilities. No further processing was performed on this ground corn. This allowed for the direct comparison of test kit results from the facilities with the instrument results from the OTSC lab, and eliminated some variability associated with sampling and sample preparation.

Criterion 1 - Regression and Criterion 2 – Paired *t*-test (matched pairs)

2016 Crop Corn. A graphical representation of the dataset pairs is shown in Fig. 1. The pairs were also identified by category based on the OTSC result. The Pearson correlation positive coefficient of determination (r = 0.87) indicates a strong positive correlation, where the results from the kit and instrument tend to increase together. This translates to 76 % of the variance in the facility result can be predicted by the OTSC result. Based on the matched pairs analysis (Table 3), the overall mean difference of −6.4 showed that the facility mean was less than the OTSC mean, and this difference was highly significant (p-value <0.0001). This could lead to an underestimation of the true contamination. The criteria set for the paired *t*-test was for no significant difference at 95 % confidence. This means that the p-value would need to be greater than 0.05. If validation was based solely on this criteria, then the test kit would not pass.

The pairs were also placed into bins based on the OTSC result. Bins included <30 ppm, ≥30 ppm ≤ 60 ppm, >60 ppm. Further assessment showed that for concentrations <30 ppm, the mean difference was +1.9 ppm, and was nearly insignificant (p-value = 0.045). However for levels ≥30 ppm, the mean difference was a large as −27.3 ppm and highly significant. The goal of the blending plan is to "dilute" contaminated corn with enough clean corn to achieve a fumonisin level below 60 ppm. The test kit was underestimating the concentration at these higher levels. It is not surprising that the UPLC-MS-MS method can better analyze for fumonisin in a dynamic concentration range. The difference for the test kit could be attributed to analyst error, the need for additional dilutions when using the test kit, and/or reaching an upper limit of quantification for the test kit.

2017 Crop Corn. A graphical representation of the dataset pairs is shown in Fig. 2. The pairs were also identified by category based on the OTSC result. Similar to the 2016 Crop Corn, the Pearson correlation coefficient of determination (r = 0.90) indicates a strong positive correlation. This translates to 82 % of the variance in the facility result can be predicted by the OTSC result. Based on the matched pairs analysis (Table 3), the overall mean difference of 10.3 ppm showed that the facility mean was greater than the OTSC mean, and this difference was highly significant (p-value <0.0001). If validation was based solely on this criteria, then the test kit would not pass.

Unlike the previous dataset for the blend plan, the mean difference shows that the facility result is higher than the OTSC result. Test kits do undergo updates and improvements. It is possible that the test kit made adjustments to decrease the risk of false negatives (by overestimating the levels). This is more apparent when further assessing the data based on concentration levels. For the >60 ppm category, the facility mean difference is now a +24.1 ppm compared to the previous dataset value of −27.3 ppm. Based on the mean difference, there also appears to be less alignment (greater mean differences) in the <30 ppm and ≥30 ppm ≤ 60 ppm bins.

3.3. *Criterion 3 – application of FGIS performance criteria for assessment*

An overall 90 % pass rate, meaning 90 % of the results meet the FGIS range criteria for a test kit, was set as the requirement for validation of test kit. The FGIS performance criteria are set at 95 %, but this is for qualification of the kit under ideal circumstances with pre-set concentrations for the materials being tested. The 90 % level would provide enough assurance that the test kit is still performing at an acceptable level. If the UPLC-MS-MS result is taken as the "true" value for the sample, then the test kit value should have a result that falls within the range based on the FGIS performance criteria. An overall 90 % pass rate would validate the test kit.

2016 Crop (216 entries related to blending plan). For this data set, only 54 % of the results fell within the calculated range for the corresponding concentration (Fig. 3). A further breakdown of the data showed that of the 46 % that were out of range, 71 % were

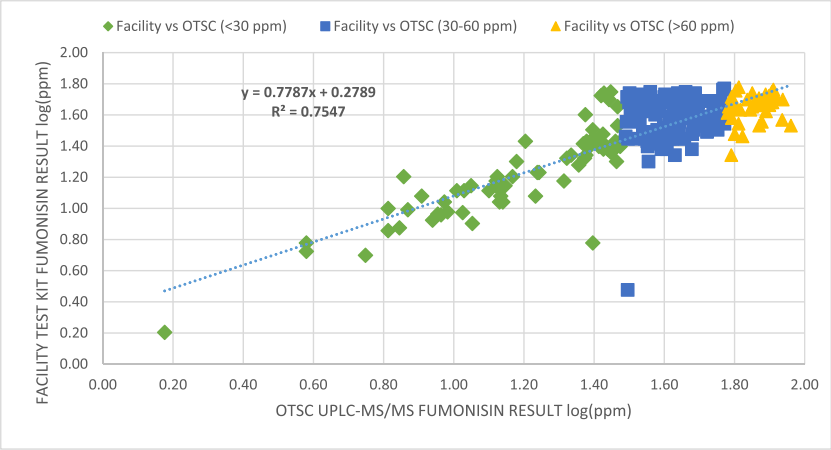


Fig. 1. Facility Kit Result vs OTSC UPLC-MS-MS Result for Dataset 1: Blend Plan Facilities (2016 Corn Blend Plan results) (ppm = mg/kg).

Table 3
Matched Pairs Evaluation: 2016 Crop Corn (blend plan verifications) and 2017 Crop Corn (OSS verifications).

Dataset	Category	Facility Mean	OTSC Mean	Mean Difference	Standard Error	P-value Prob > |t|	N =
2016 Crop Corn[a]	Overall	33.8	40.2	−6.4	0.9885	<0.0001	216
	<30 ppm	19.4	17.5	1.9	0.9380	0.0445	69
	≥30 ppm ≤ 60 ppm	38.9	43.0	−4.1	1.131	0.0004	108
	>60 ppm	45.3	72.6	−27.3	1.674	<0.0001	39
2017 Crop Corn[b]	Overall	42.0	31.7	10.3	1.294	<0.0001	177
	<30 ppm	19.8	13.2	6.6	0.6556	<0.0001	115
	≥30 ppm ≤ 60 ppm	52.1	40.9	11.2	1.848	<0.0001	34
	>60 ppm	121.0	96.9	24.1	6.876	0.0016	28

[a] 2016 Crop Corn (blend plan) includes 6 facilities.
[b] 2017 Crop Corn (OSS verifications) includes 7 facilities.

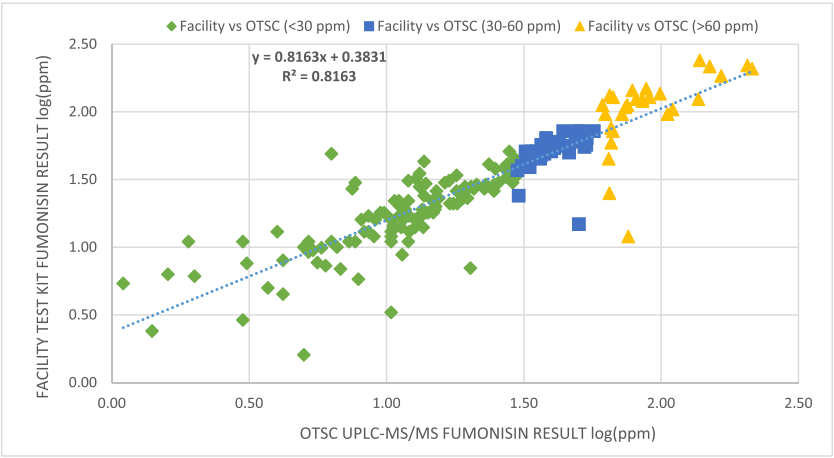


Fig. 2. Facility Kit Result vs OTSC UPLC-MS-MS Result for Dataset 2 - Seven OSS Facilities (2017 Verification results) (ppm = mg/kg).

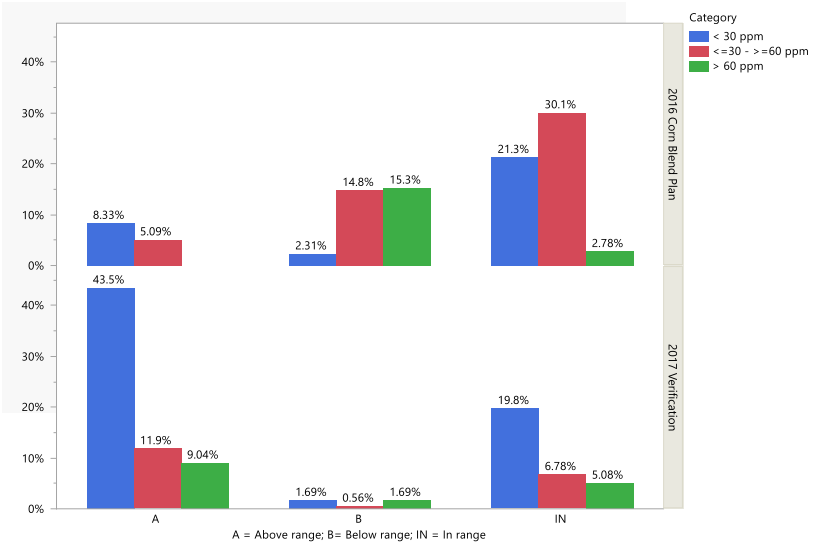


Fig. 3. Application of FGIS Criteria for Range: 2016 Corn Blend Plan and 2017 Verifications. Percentages are based on grand total within each data set (2016 crop – blend plan samples, N = 216; and 2017 crop - verification samples, N = 177) (ppm = mg/kg).

below the range. An additional assessment based on concentration categorization, <30 ppm, ≥30 ppm ≤ 60 ppm, and >60 ppm showed a similar trend for results ≥30 ppm with a greater percentage falling below the range. For <30 ppm, 67 % of the test kit results fell within range. If the determined risk is lower by having the kit result above (high bias), then for the <30 ppm category, the kit would pass with 93 %. For the >60 ppm category, the test kit failed to meet the range requirements 83 % of the time, with 100 % of those results "Below range". This is by far the greater risk – for a kit to underestimate the fumonisin concentration in the sample. Correct classification based on concentration is especially critical on decisions related to the corn's usage.

2017 Crop (177 verification samples from 7 OSS facilities). For this data set, only 32 % of the results fell within the calculated range for the corresponding concentration (Fig. 3). A further breakdown of the data showed that of the 68 % that were out of range, 94 % were above the range. An additional assessment based on concentration categorization, <30 ppm, ≥30 ppm ≤ 60 ppm, and >60 ppm showed a similar trend for results in all categories with a greater percentage falling above the range. If the determined risk is lower by having the kit result above (high bias), then overall and for all categories, the kit would pass with >90 %. This shows the kit mitigates the risk of a false negative with an overestimation of the fumonisin concentration in the sample. This is opposite of what was seen for the 2016 Crop data. For the 2016 crop, a majority of the results were "below range" whereas for the 2017 crop, a majority were "above range". A change in the test kit between these years may account for these differences. Though an overestimation of the fumonisin level is a lower risk related to health of an animal, it may create hardships for the producer due to discounts on the corn.

Criterion 4 – Use of violation status: Violation set at > 5 ppm or >60 ppm.

2016 Crop. Of the 216 result pairs (Facility vs OTSC), there was 98.6 % agreement on test results (combination of agreement with non-violative and violative samples) (Table 4). 1.4 % of the samples were incorrectly designated as non-violative or violative by the facility. Of this subset, only 0.5 % (or 1 sample out of 216) was labeled as non-violative when it should have been violative. If using Criterion 4 for assessment, the test kit passes validation by meeting the 90 % agreement (combined agreement on classification of non-violative and violative) and a Type II error of less than 1 %.

Blend plans were initiated with regulatory results >60 ppm. If the violation status is set to this condition, the test kit only achieves 82 % agreement (Table 4). Because these particular data were from blend plans, the results should have been all less than 60 ppm. The facility results were all less than 60 ppm, but the OTSC result for the same samples showed otherwise. In turn, the Type II error was nearly 18 %.

2017 Crop. Of the 177 result pairs (Facility vs OTSC), there was 93.3 % agreement on test results (combination of agreement with non-violative and violative samples) (Table 4). 6.7 % of the samples were incorrectly designated as non-violative or violative by the facility. Of this subset, 1.1 % (or 2 samples out of 177) were labeled as non-violative when they should have been violative. If using Criterion 4 for assessment, the test kit passes validation by meeting the 90 % agreement (combined agreement on classification of non-violative and violative). The Type II error of less than 1 % was almost met. Adjusting the violation limit to >60 ppm, the agreement on test results remains above 90 % (Table 4). Though the percentage of incorrectly designated samples was the same as the <5 ppm, there was a greater percentage of false-negatives.

4. Conclusion

A summary of the test kit validation performance metrics is provided in Table 5. The quality measures were met for both datasets and provided confidence that particle size was consistent and that the test kit was performing accordingly. When the test kit was used to measure reference material (working controls) at different sites by different operators, the results met specifications. The quality measures allow for the retained samples to be representative of the particular load of corn. The OSS program has previously been shown to improve the consistency in testing through its training and qualification of operators at the facilities [22]. Verification of the result from the retained sample by OTSC should closely mirror the facility's result. Regression analysis (Criterion 1) provided information of the relationship between the facility and OTSC results. It was expected that this relationship be positive. The blend plan dataset and the OSS facility dataset failed the acceptance criteria for the paired *t*-test (Criterion 2). However, it is to be expected that a test kit would perform differently than a high-end instrument like a UPLC-MS-MS. In the first dataset, the overall mean difference showed the test kit underestimating the result. In the second dataset, the overall mean difference showed the test kit overestimating the

Table 4
Crop Year Violation Status: Rate of agreement of test results.

	Crop Year 2016		Crop Year 2017	
	Facility Non-violative	Facility Violative	Facility Non-violative	Facility Violative
Violation > 5 ppm				
OTSC Non-violative	1.4 %	0.9 %	2.3 %	5.6 %
OTSC Violative	0.5 %	97.2 %	1.1 %	91.0 %
Violation > 60 ppm				
OTSC Non-violative	81.9 %	0.0 %	79.7 %	4.5 %
OTSC Violative	17.6 %	0.5 %	2.3 %	13.5 %

Table 5
Summary of test kit validation performance metrics.

	Dataset 1 2016 Crop (Blend Plan) Pass/Fail	Dataset 2 2017 Crop (OSS facilities) Pass/Fail
Quality Measures		
Working Controls	Pass	Pass
Particle Size	Pass	Pass
Data Analysis Acceptance Criteria		
Criterion 1 – Regression	Positive, r = 0.87	Positive, r = 0.90
Criterion 2 – Paired *t*-test (overall)	Fail, p-value <0.5	Fail, p-value <0.5
Criterion 3 - FGIS Performance Criteria range	Fail	Pass[a]
Criterion 4 – Violation Classification >5 ppm	Pass	Pass[b]
Criterion 4 – Violation Classification >60 ppm	Fail	Fail[c]

[a] Pass if criteria includes those points "Above range", which would be less of a risk.
[b] Type II error was slightly above 1 %.
[c] Agreement >90 % but Type II error was 2.3 %.

result. In the testing period between these two datasets, it is possible that the test kit underwent some adjustments as well. At the time of the study, conversations with Charm Science were on-going. The Service was working in collaboration with the company to help align their kits with the reference material being used by the OSS facilities. As a result of this collaboration, the company did make a change to the source of their fumonisin certified analytical fumonisin standards, opting to use those made by Bio-pure Romer. Training on usage of the kit may have been better for the second dataset since these facilities were actively part of the OSS program.

For FGIS Performance Criteria Range (Criterion 3), dataset 1 (2016 crop) did not meet the 90 % pass rate even if the criteria was modified to include those points "Above range". However, dataset 2 (2017 crop) does meet the 90 % pass rate with this modification. If risk is factored into the criteria, including those points that were "Above range" would provide a conservative approach to the validation. For Violation Classification (Criterion 4), in both datasets, over 90 % of the samples were classified correctly when the violation classification was set to > 5 ppm. Dataset 1 had less than 1 % of its samples classified incorrectly as violative when they were non-violative. Dataset 2 had a greater percentage of its samples classified incorrectly as violative when they were non-violative (5.6 %). However, the test kit appeared to manage Type II error in both datasets with rates of 0.5 % and 1.1 %.

Each of the criterion for data analysis acceptance provided different information on kit performance, and a combination may be the best solution for field validation. The proposed scheme for field validation of a test kit would need to include quality measures of working controls using matrix-matched reference materials, particle size analysis of the ground corn, and analyst training and qualification (Fig. 4). Working controls (those run on the testing days) and verification results (confirmation of results on retained material) would need to be acceptable based on specified criteria. Of this criteria, most importantly, the test kit should correctly classify the corn as violative or non-violative based on determined regulatory guidelines.

Data availability statement

Data will be made available upon request.

CRediT authorship contribution statement

Megan K. Rooney: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Timothy J. Herrman:** Writing – review & editing, Supervision, Resources, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used OpenAI's ChatGPT 4.0 in order to improve language and readability. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

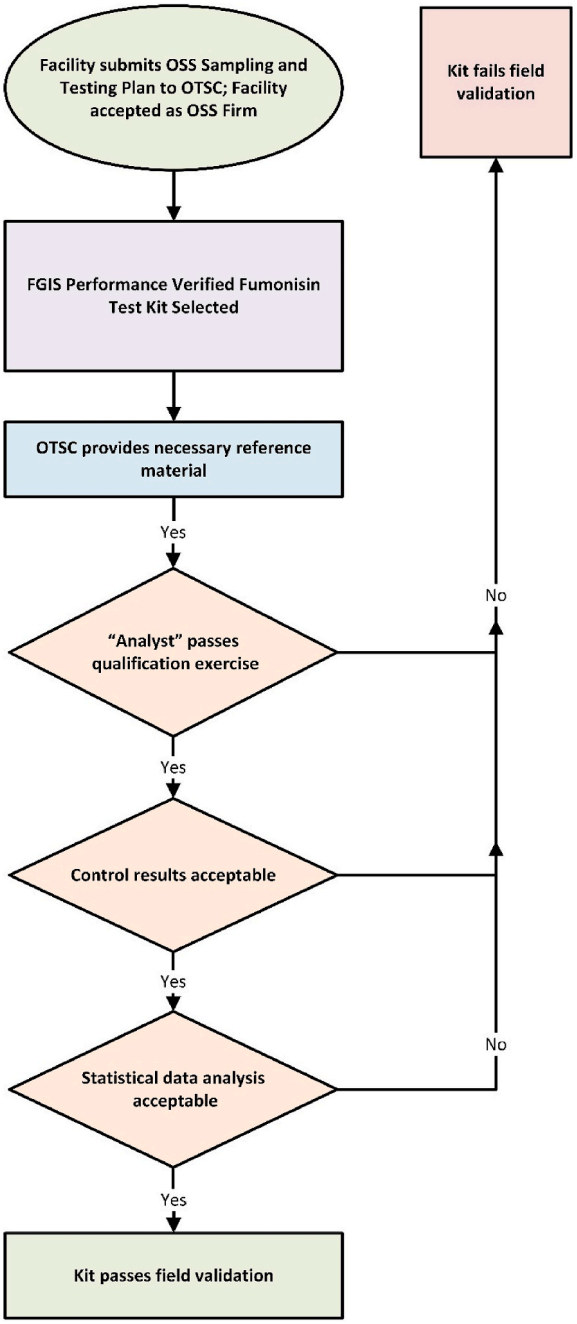


Fig. 4. Proposed scheme for field validation.

Appendices.

Appendix 1
Quality Measures/Record Review – 2016 Corn Crop - Blend Plan Facilities

Facility	# of Controls run*	% of controls within range	Particle Size Records N =	Particle Size (>70 % fines) Mean =
Facility A	3	100 %	2	94

(*continued on next page*)

Appendix 1 (*continued*)

Facility	# of Controls run*	% of controls within range	Particle Size Records N =	Particle Size (>70 % fines) Mean =
Facility B	2	100 %	1	91
Facility C	9	100 %	7	80
Facility D	4	100 %	18	88
Facility E	5	100 %	4	82
Facility F	2	100 %	2	88

* 47 ppm control: Acceptable range of 33 ppm–61 ppm; ±30 % given due to extra dilution required for analysis). Source for Quality Measures [20].

Appendix 2

Quality Measures/Record Review – 2017 Corn Crop - from Qualified Facilities for OSS Program

Facility	# of controls run	At least 20 controls run?	% of controls within the ± 20 % range	% of controls within range adjusting for a 30 % range for 63 ppm control*	Particle Size Records N =	Particle Size >70 % fines Mean =
Facility 1	51	Yes	41/51 80.3 %	51/51 100 %	57	87 %
Facility 2	30	Yes	29/30 96.7 %	29/30 96.7 %	17	91 %
Facility 3	45	Yes	45/45 100 %	45/45 100 %	23	84 %
Facility 4	39	Yes	38/39 97.4 %	39/39 100 %	44	83 %
Facility 5	31	Yes	29/31 93.5 %	31/31 100 %	33	81 %
Facility 6	32	Yes	29/32 90.6 %	32/32 100 %	24	82 %
Facility 7	31	Yes	31/31 100 %	31/31 100 %	10	88 %
Overall	**259**	**Yes**	**242/259** **93.4 %**	**258/259** **99.6 %**	**208**	**85 %**

*Note: At the time of this study, the acceptable range was set at 20 % for the OSS program. The acceptable range for fumonisin is now set at ±30 % for all concentration levels. Source for Quality Measures [20].

Glossary

AOAC	Association of Official Analytical Chemists
EU	European Union
FDA	Federal Drug Administration
FGIS	Federal Grain Inspection Service
ISO	International Organization for Standardization
ISO/IEC	International Organization for Standardization/International Electrotechnical Commission
OSS	One Sample Strategy
OTSC	Office of the Texas State Chemist
ppb	Parts per billion (μg/kg)
ppm	Parts per million (mg/kg)
RSD	Relative Standard Deviation
SD	Standard Deviation
TAC	Texas Administrative Code
TX	Texas
UPLC-MS-MS	Ultra-high performance liquid chromatography tandem mass spectrometry
US	United States
USDA	United States Department of Agriculture
USDA-AMS-FGIS	United States Department of Agriculture - Agricultural Marketing Service – Federal Grain Inspection Service

References

[1] S.Y. Dai, et al., Aflatoxin risk management in Texas: test kit approval for maize, Journal of Regulatory Science 1 (1) (2013) 15–22.

[2] United States Department of Agriculture (USDA), Agriculture Marketing Service (AMS), and Federal Grain Inspection Service (FGIS) *Design Criteria and Test Performance Specifications for Quantitative Fumonisin Test Kits*, 2018.

[3] Texas Administrative Code (TAC), *Title 4*. Agriculture Chapter 61 Commerical Feed Rules, Office of the Texas State Chemist (OTSC), 2019.

[4] Federal Grain Inspection Service (FGIS), FGIS approved mycotoxin rapid test kits 08-30-2022, Retrived from: https://www.ams.usda.gov/sites/default/files/media/FGISApprovedMycotoxinRapidTestKits.pdf, 2022. (Accessed 18 October 2023).

[5] W. Li, *CHARM FUMO WETS5 Validation*, Office of the Texas State Chemist, 2016. (Accessed 29 November 2016).

[6] D. Pickova, et al., Aflatoxins: history, significant milestones, recent data on their toxicity and ways to mitigation, Toxins 13 (6) (2021).

[7] Food and Drug Administration (FDA), Guidance for Industry: Fumonisin Levels in Human Foods and Animal Feeds, vol. 2001, 2001. Retrieved from, https://www.fda.gov/regulatory-information/search-fda-guidance-documents/guidance-industry-fumonisin-levels-human-foods-and-animal-feeds. (Accessed 18 October 2023).
[8] Texas Administrative Code (TAC), *Title 4. Agriculture*. Chapter 61 - Commercial Feed Rules (Amended May 19, 2011), 2011.
[9] T.J. Bucci, P.C. Howard, Effect of fumonisin mycotoxins in animals, J. Toxicol. - Toxin Rev. 15 (3) (1996) 293–302.
[10] W.F.O. Marasas, et al., Fumonisins disrupt sphingolipid metabolism, folate transport, and neural tube development in embryo culture and in vivo: a potential risk factor for human neural tube defects among populations consuming fumonisin-contaminated maize, J. Nutr. 134 (4) (2004) 711–716.
[11] A.H. Merrill, et al., Sphingolipid metabolism: roles in signal transduction and disruption by fumonisins, Environmental Health Perspectives 109 (2001) 283–289.
[12] K.R.N. Reddy, et al., An overview of mycotoxin contamination in foods and its implications for human health, Toxin Rev. 29 (1) (2010) 3–26.
[13] American Association for Laboratory Accreditation (A2LA), A2LA directory of accredited organizations, 2023, https://customer.a2la.org/index.cfm?event=directory.index, 2023. (Accessed 18 October 2023).
[14] W. Li, T.J. Herrman, S.Y. Dai, Rapid determination of fumonisins in corn-based products by liquid chromatography/tandem mass spectrometry, J. AOAC Int. 93 (5) (2010) 1472–1481.
[15] Association of Official Analytical Chemists (AOAC), Appendix F: Guidelines for Standard Method Performance Requirements, AOAC International, Gaithersburg, MD, 2016.
[16] W. Horwitz, R. Albert, The Horwitz ratio (horrat): a useful index of method performance with respect to precision, J. AOAC Int. 89 (2006) 1095–1109.
[17] International Organization for Standardization (ISO), ISO 13528:2022(E) Statistical methods for use in proficiency testing laboratory comparison, in: Annex B: Homogeneity and Stability of Proficiency Test Items, 2022.
[18] United States Department of Agriculture (USDA), Agriculture Marketing Service (AMS), and Federal Grain Inspection Service (FGIS) *Test Kit Instruction: Charm ROSA WET-S5 Fumonisin Quantitative Test Using Charm EZ-M Reader*, 2022.
[19] Charm Sciences, ROSA WET-S5 Fumonisin Quantitative Test Flow Chart (3:1) - High PPM, 2016, Charm Sciences, Inc., 2016.
[20] Office of the Texas State Chemist (OTSC), One sample Strategy for mycotoxin risk management in Texas - 2023 Handbook. https://otsc.qualtraxcloud.com/Showdocument.aspx?ID=7186, 2023.
[21] W. Li. Validation of Fumonisin in feed SOP 16711 LC/MS/MS, Office of the Texas State Chemist, 2009 (Accessed 05 April 2024). Available from: Office of the Texas State Chemist, College Station, TX; IQM Doc ID: 15165.
[22] M. Sasser, T.J. Herrman, K.M. Lee, Evaluation of coregulation as a governance option to manage aflatoxin risk in Texas maize, J. Food Protect. 81 (4) (2018) 554–560.