Food Chemistry 528 (2026) 151037

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

Monochromatic 'high-definition' XRF is a highly robust tool for high-throughput lead screening of foods

Hala Baidas, Melikizat Begzad, Sadia Muneem, Sinisa Urban *

Division of Environmental Sciences, Laboratories Administration, Maryland Department of Health, 1770 Ashland Avenue, Baltimore 21205, MD, USA

ARTICLE INFO

Keywords:
Lead
X-ray fluorescence
Screening method
Validation
Closer to zero
Spices

ABSTRACT

Spices are well-known sources of harmful lead exposure, yet approved food testing methods, while sensitive and robust, are nevertheless slow, expensive, and labor-intensive. We addressed the need for a lead screening method in spices by leveraging monochromatic 'high-definition' X-ray fluorescence (mHD-XRF). By strictly applying FDA chemical method validation guidelines with 136 diverse spices collected at retail by 3 states over 13 years, we established mHD-XRF as an ideal screening tool. The level of detection was 0.044 ppm, which is ~100-fold more sensitive than conventional XRF, and quantitation remained within 20% of 'gold-standard' values across >3 orders-of-magnitude. We set the screening threshold at 1.0 ppm, and encountered no false-negatives or false-positives in all 136 spices. FDA experts reviewed our level-2 single laboratory validation favorably, and the method was successfully added to our ISO/IEC-17025:2017 accreditation. Widespread operationalization of this method, especially at ports-of-entry, promises to increase spice surveillance by orders of magnitude.

1. Introduction

There is no safe level of lead, which makes prevention of chronic exposure paramount (Centers for Disease Control and Prevention, 2025a). Lead bioaccumulates in humans, such that even small exposures continuously add-up over time, and exerts detrimental and irreversible damage to multiple systems including the brain and kidneys at low levels. As such, lowering exposures is a shared goal of multiple public health and environmental programs. In fact, in 2021 CDC reduced the blood lead reference value from 5.0 to 3.5 μg/dL (Centers for Disease Control and Prevention, 2025b), further mandating the need to reduce even low-level chronic exposures. This has become an escalating challenge: as actions are taken to reduce higher exposures, previously uncommon or low exposures take on a greater share of a person's body burden.

EPA estimates that approximately half of a person's exposure to environmental lead is through food ingestion (Environmental Protection Agency, 2017; Neltner, 2017), and spices are known to be a major source of lead exposure from foods (Ishida et al., 2022). This results from the fact that plants readily take-up and concentrate environmental lead in their tissues, which become the source of commonly used spices (Shahid et al., 2017). Coupled with the fact that many spices are produced in varied parts of the world where there may not be effective environmental controls, spices are a persistent public health vulnerability. Moreover, selling of 'off brand' spices in discount stores likely constitutes a growing environmental justice issue, because disadvantaged communities are likely to receive lower levels of protection than populations that shop at more expensive stores.

FDA protects the public's health by removing foods that pose health risks including from naturally occurring toxic elements like lead. The 'gold standard' for laboratory testing are methods like FDA's Elemental Analytical Manual (EAM) 4.7 that use microwave-assisted acid digestion to free lead ions from foods, and measure their abundance with tremendous sensitivity using inductively coupled plasma mass spectrometry (ICP-MS) (Gray, Mindak, Cheng, & Kovalenko, 2025). Low and even sub parts-per-billion (ppb) levels of quantification can routinely be achieved. However, these methods rely on sophisticated and expensive instrumentation, require significant operator training and experience, remain slow and labor intensive, and incur significant reagent costs to run.

Maryland's public health laboratory, as one instructive example, routinely runs no more than 6 food samples per week for toxic metals testing using EAM 4.7, and takes at least 4 days of work to arrive at a lead measurement. Therefore, while surveillance testing of spices available in grocery stores stands as the last line of defense for protecting the public from harmful lead exposures, the scale of this testing remains

* Corresponding author.
E-mail address: sinisa.urban@maryland.gov (S. Urban).

https://doi.org/10.1016/j.foodchem.2026.151037
Received 12 April 2026; Received in revised form 12 August 2026; Accepted 2 September 2026
Available online 8 September 2026

limited and unable to meet public health program needs.

These testing challenges are not unique to lead; rapid and sensitive screening methodologies have been developed for microbiological threats that overcome these burdens (Food and Drug Administration, 2026a). These screening methods are first able to detect specific microbial DNA or protein, and if detected, testing reflexes to lengthier confirmation techniques that culture and isolate an isogenic pathogen. These screening methodologies have allowed surveillance testing for microbial threats to be increased by orders of magnitude. However, despite their proven success, screening technologies are rare, if not entirely absent, in the realm of chemical threats.

In addition to ICP-MS, X-ray fluorescence (XRF) is able to detect toxic elements including lead in various matrices: incident X-rays impart energy that ejects inner-shell electrons from atoms, and as outer-shell electrons take their place, X-rays of a lower energy are emitted and quantified (Marguí, Queralt, & de Almeida, 2022; Revenko & Pashkova, 2023). This non-destructive spectroscopic technique requires minimal sample preparation and is thus inexpensive and rapid, but it suffers from low sensitivity, partly because a wide spectrum of X-rays is used to illuminate the sample, making distinguishing true emitted fluorescence at different wavelengths from unavoidable multi-wavelength scatter difficult. This inherent limitation makes it challenging to deconvolute the spectra and quantify the elements present.

More recently a new generation of XRF analyzers have been developed that use a doubly curved crystal (DCC) to focus polychromatic incident X-rays into a monochromatic beam before they illuminate a sample (Chen & Gibson, 2002; Chen, Gibson, & Huang, 2008; Chen & Wittry, 1998). The dramatic reduction in spectral complexity and background, coupled with diligent background processing using a Fundamental Parameters algorithm (de Boer, 1989; Kitov, 2000), have resulted in a ~ 100-fold increase in sensitivity for toxic elements like lead.

While previous studies have demonstrated high sensitivity of these analyzers with lead (Johnson-Restrepo, Blain, Judd, Tysoe, & Parsons, 2025) and cadmium (Cifuentes, Rodriguez, Avellaneda-Torres, Quiroga-Mateus, & Bravo, 2026) in certain foods, they have never been evaluated for their performance in criteria set forth for regulatory food screening programs by the FDA (Food and Drug Administration, 2019). Since compliance testing labs cannot use an unvalidated method even despite its established scientific promise, we sought to evaluate mHD-XRF specifically for regulatory testing. Since no method existed, our work described here constitutes a first-of-its-kind single laboratory validation study, and therefore necessarily included a direct comparison to the FDA approved method EAM 4.7 that uses microwave-assisted digestion and ICP-MS.

2. Materials and methods

2.1. Sample collection, compositing, and homogenization

Since assembling a large sample set of real-world retail spices of at least 59 samples above 1 ppm lead was critical but beyond the experience of a single laboratory, samples from the New York State Department of Agriculture and Markets (n = 49) and the Connecticut Agricultural Experiment Station (n = 12) were generously provided and added to the Maryland Department of Health sample set (n = 75). Importantly, all spice samples were collected for regulatory testing from retail outlets by trained collectors for both state and federal programs in Maryland, New York, Connecticut over a 13-year period (2014–2026). For the FDA Toxic Elements Compliance Program (Food and Drug Administration, 2026b), this necessitated collecting 12 retail units (subsamples) of the same lot.

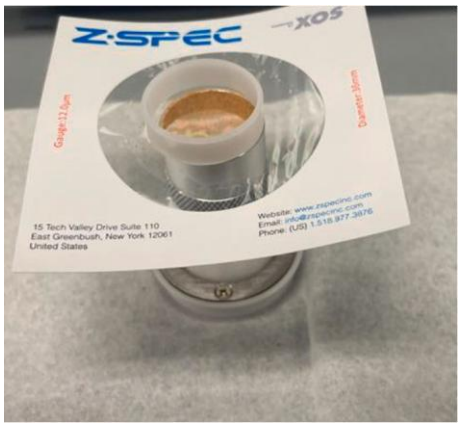
An analytical balance (0.1 mg sensitivity; Mettler Toledo) was used to weigh equal amounts (120 g / 4 oz) of each subsample, as per the FDA Compliance Program requirements, which were combined to make a gravimetric composite of the 12 subsamples (if the total subsample mass was less than 4 oz. the composite was made using the full quantity of each subsample). All sample handling and preparation were performed in polypropylene metal-free hoods to minimize potential trace metal contamination. Disposable polypropylene smartSpatulas (LevGo, Cat. No. 17221) were used for sample handling to minimize potential trace metal contamination. The composite of the subsamples was transferred to a freshly acid-washed food processor (Waring 3.5-Quart Food Processor with Flat Lid Model WFP14SW). Alternatively, when only a single retail unit was analyzed, 4 oz., or if less than 4 oz., then the entire product or the amount of product to the maximum fill line, was transferred to a coffee grinder (SP SCIENCEWARE Mill Grinder F37250–0010, or Hamilton Beach coffee grinder). Food processors and grinders were acid-washed before each use, and periodically verified to be lead-free. Samples were ground for ~2 min at room temperature, which was selected to ensure that the resulting material was visually uniform in particle size.

2.2. Spice analysis by mHD-XRF

Homogenized spice samples were scanned in a Zmax mHD-XRF analyzer (*Z*-Spec Inc., operating software version 2.1.4-47d70c3 and firmware version EMAX-ST-2.0.0). Optimization of the Fundamental Parameters model for the Zmax was initially performed by the manufacturer using 8 NIST Standard Reference Material and European Reference Materials (SRM1568c, rice, 0.0168 ppm; SRM1577c, bovine liver, 0.0628 ppm; ERM BD151, milk powder, 0.207 ppm; SRM1570, spinach leaves, 0.2 ppm, SRM1566b, oyster tissue, 0.308 ppm; SRM1515, apple leaves, 0.47 ppm; SRM1547, peach leaves, 0.869 ppm). An additional NIST SRM (mussel tissue, NIST SRM 2976, 1.19 ppm) and a cinnamon consumer product previously tested by ICP-MS (4.3 ppm) at the Maryland Department of Health Laboratories Administration were added to allow further refinement of the Fundamental Parameters model by the manufacturer. Instrument performance was verified with standards and proved to be $R^2 = 0.9993$.

Samples were prepared by creating a thin layer of ground spice between disposable polypropylene two films (12.0 μm, *Z*-Spec, part # 200335–0). A first film was laid over the opening of a sample pedestal/cup (Z-Spec, part # 309006–01). A sample film metal ring (Z-Spec, part # 304770–02) was placed over the film and pushed down so that the film was captured under the ring and the ring was secured around the opening of the sample pedestal/cup:

Once the ring was fitted to the sample pedestal/cup, the cardboard support and excess film were removed. The sample pedestal/cup was placed film side up on an analytical balance. Approximately 0.08 to 0.09 g of sample was measured onto the top of the film. After the required weight was obtained, the sample pedestal/cup was removed from the scale and another polypropylene film (12.0 μm, *Z*-Spec, part # 200335–0) was placed on top of the sample:



The sample film plastic ring (O-ring Z-Spec, part # 204662–01) was pushed over the second film, and slid down the sample pedestal/cup so that the second film was captured under the ring:

A thin layer of sample was thereby secured between the two films. Once the plastic ring was pushed below the top of the metal ring, the cardboard support and excess film were removed. A clean tongue depressor was used to smooth the sample under the film to ensure that it was spread evenly and fully covered the film.

Prior to beginning sample analysis each day, a Quality Control sample was tested to verify unit performance. In this case, NIST 2976 (freeze-dried mussel tissue, NIST, Gaithersburg) or NIST 3299 (ground turmeric, NIST, Gaithersburg) were used as the true value, since 1.19 ppm and 1.20 ppm were close to the chosen trigger/threshold level of 1 ppm. All Quality Control samples tested during the instrument validation measured within 20% of the true value.

Once the sample pedestal/cup was placed into the guide, the sliding door was closed, and the instrument parameters were set. Each sample was tested with material set as powder, test time at 600 s, reading refresh intervals of 30 s, number of Samples set to 1, and calibration as default. All samples were tested under these same parameters.

2.3. Determination of mHD-XRF detection and quantification limits

The LOD and LOQ of mHD-XRF were calculated based on testing 9 replicates of a spice containing a low lead level of 0.058 ppm:

$$LOD = 3 \times SD$$

$$LOQ = 10 \times SD$$

where S is the standard deviation of the 9 replicate measurements.

An accuracy and precision study was performed by testing 9 replicates of spices at 4 levels: blank (rice), low level (below the trigger level), medium (around the trigger level), and high (above the trigger level) at 0.0, 0.624, 1.42, and 2.74 ppm, respectively. The standard deviation and relative standard deviation were calculated as well as the recovery based on the mean of the 9 replicates.

2.4. ICP-MS analysis

All spices were also previously or concurrently tested by ICP-MS in duplicate. Quartz digestion vessels and caps (SCP Science, Part No. 010–600-071-N) were soaked overnight in 10% HNO_3 (*v*/v, Veritas double-distilled nitric acid, GFS Chemicals, Part No. 621). Before use, fresh 10 mL of 10% (*v*/v) HNO_3 was added to each vessel and prepared using rinse program conditions that were the same as digestion parameters except with a hold time extended to 25 min. All labware was thoroughly rinsed with reagent water and allowed to dry completely before use.

An analytical portion (0.5 g) of each composited sample was weighed directly into each of two 75 mL quartz digestion vessels. Reagent water (1 mL, purified using a Millipore Quantum ICP cartridge system and with a resistivity of ≥18.0 MΩ·cm at 25 °C), concentrated HNO_3 (8.0 mL, 67–70%, plasma purity grade, SCP Science, Part No. 250–036-135), and H_2O_2 (1.0 mL, 30% trace metals grade, Avantor, CAS No. 7722-84-1) were added to each vessel, and the vessels were capped with Teflon caps. Samples were digested using the NovaWAVE FA microwave digestion system fitted with independent pressure sensors and temperature monitoring for each vessel (SCP Science, Canada). The system was programmed to ramp to 200 °C and maintain pressures up to 500 psi to obtain complete matrix decomposition. Microwave digestion conditions used were microwave power of 12 × 250 W, 2.45 GHz, and ramp and hold times of 15 min each.

The reagent blank and rinse solution with 1% (*v*/v) HNO_3 + 0.5% HCl and containing 100 μg/kg Au was prepared by adding 10 mL concentrated HNO_3, 5 mL HCl and 0.1 mL Au stock solution (1000 μg/mL in 2% HCl; High Purity Standards, Part No. 100021–2) to approximately 800 mL reagent water and diluting to 1 L. This solution was used as the calibration blank (Cal 0), rinse solution, diluent for digested samples, and for continuing calibration blanks (CCB). The carrier solution contained of 2% (*v*/v) HNO_3, prepared by diluting 20 mL concentrated HNO_3 to 1 L with reagent water. Internal standard stock solutions included a multi-element internal standard solution containing Bi, Ho, In, Li, Sc, Tb, and Y (10 μg/mL each in 2% HNO_3; Spex CertiPrep, Part No. CLISS-1) and a germanium internal standard solution (1000 μg/mL in 2% HNO_3; High Purity Standards, Part No. 100020–3). The internal standard intermediate solution was prepared in a 1 L volumetric flask by adding 50 mL Veritas double-distilled HNO_3, 40 mL isopropanol, 0.5 mL Ge internal standard stock solution, and 10 mL multi-element internal standard stock solution, followed by dilution to 1 L with reagent water.

Method blanks (MBK) were prepared by adding 0.5 mL reagent water to digestion vessels and processing them in the same procedure as the samples. Reference materials were weighed as 0.5 g portions and digested under the same procedure. DOLT-5 (dogfish liver) was purchased from the National Research Council of Canada (NRC, Canada), and SRM 2976 (mussel tissue) was purchased from the National Institute of Standards and Technology (NIST, Gaithersburg, MD, USA). Fortified method blanks (FMB) were prepared by adding 0.5 mL reagent water to digestion vessels followed by fortification with the second source of spiking standard prior to digestion. Fortified analytical portions (FAP) were prepared by weighing 0.5 g of sample and fortifying with the second source spiking standard before digestion. Disposable graduated transfer pipets (Fisher Scientific, Cat. No. 13–711-9 AM) were used for sample and standard preparation. Certified disposable 50 mL test tubes

(SCP Science, Part No. 010–500-263) were used for analytical solutions. Chemical-resistant, disposable, powder-free gloves were worn during sample handling and digestion procedures to ensure laboratory safety and prevent contamination.

After digestion, vessel contents were transferred to a 50 mL centrifuge tube, vessel walls were rinsed three times with reagent blank solution and added to the centrifuge tube, and the analytical solution was brought to a final mass of 50 *g*. The solutions were centrifuged for 25 min at 3774*g* (Thermo Scientific, Sorvall ST4 Plus-MD), or allowed to settle overnight prior to analysis.

Digests were analyzed for lead using a Perkin-Elmer NexION 350 ICP-MS (Perkin-Elmer, Waltham, MA, USA) run in kinetic energy discrimination (KED) mode with helium collision gas at a flow rate of 4.5 mL/min to reduce polyatomic interferences. Liquid argon (99.99% purity) was used as the plasma gas, and helium (99.9999% purity) was used as the collision cell gas. A peristaltic pump was used for delivery of samples and internal standards (ISTD). Black/black peristaltic pump tubing (0.76 mm i.d.; Perkin-Elmer, Part No. N8145202) was used for both channels to maintain a 1:1 sample-to-ISTD ratio, to minimize instrumental drift. Grey/Gy drain tubing (1.30 mm i.d.; Perkin-Elmer, Part No. N8145173) was used for waste removal. An ICP-MS daily performance solution (Perkin-Elmer, Part No. N8145051) was used for verifying instrument performance before analysis. Calibration standards were prepared from a multi-element stock standard solution (Part No. ICP-200.8-1-100) using a reagent blank as the diluent. Calibration standard concentrations were 0.1, 1, 2, 10, 20, 50 ppb, and lead concentration was reported as $^{208}Pb + {}^{207}Pb + {}^{206}Pb$ and as an average of the two duplicates.

2.5. Determination of detection and quantification limits for ICP-MS

Analytical solution detection limit (ASDL) and analytical solution quantitation limit (ASQL) were determined from replicate method blanks ($n = 12$) according to the equations described in EAM Section 3.2:

$$ASDL = 2 \times t_{95} \times S \times \sqrt{1 + \frac{1}{n}}$$

where n is the number of blank measurements, t_{95} is the one-sided Student's t value at the 95% confidence level, and S is the standard deviation of the blank measurements.

The ASQL was defined as 10 times the standard deviation of the blanks. The method detection limit (MLOD) and method quantitation limit (MLOQ) were calculated from the ASDL and ASQL values, respectively, by multiplying by the digestive dilution factor, which was calculated as the ratio of the final mass of the analytical solution (M) to the mass of the analytical portion (m). For this study, the ASDL was 0.104 ppb, the ASQL was 0.279 ppb, the MLOD was 10.4 ppb, and the MLOQ was 27.9 ppb for lead.

2.6. Data visualization

Data was analyzed and graphed in Prism 11.0.2 (GraphPad). Paired Student's *t*-tests were selected to assess the effect of grinding as a measure of method ruggedness (see Section 3.5 and Fig. 6) because means of two related groups were being compared (same spice with or without grinding). Each spice was treated individually because each is a unique and unrelated spice type, brand, and/or manufacturing lot.

3. Results

3.1. Establishing a simple but reliable testing workflow

Screening methods are meant to assess a large number of samples for specific threats rapidly in order to identify the few samples that might require additional testing to confirm and/or quantify the presence of an adulterant. In essence, they are mean to *exclude* unadulterated samples from additional resource-intensive scrutiny, while accurately *selecting* those that are likely to require additional testing. Therefore, to be useful, they must be rapid, sensitive, and inexpensive on the one hand, and reliable on the other. These performance criteria must be established through a rigorous validation process if they are to be used for regulatory purposes.

As a starting point, we worked-out a simple workflow to support our goal of large-scale screening of spices for lead using mHD-XRF. The FDA Compliance program requires compositing of 12 retail subsamples of the same lot (Food and Drug Administration, 2026b; Gray et al., 2025), so our approach was to composite 4 oz. of each subsample, when available. These were weighed gravimetrically and composited into a food processor for grinding to a fine powder, since some spices are coarse and XRF is a surface technique. For other purposes, including for state programs and/or when only single retail units are available, we tested homogenizing individual retail units in smaller coffee grinders. In both cases commercial processors/grinders were used with the highest available grade of stainless-steel blades, tested for background lead before use, and periodically used to grind blank spice (rice) in order to verify that no lead was being carried over.

Once a uniform fine powder was achieved, ~80 mg was weighed and placed onto the mHD-XRF pedestal between 2 layers of disposable optically optimized film obtained from the manufacturer. ~80 mg was recommended by the manufacturer because it forms a thin layer that exposes maximal surface of each sample without introducing significant X-ray scattering.

Samples were loaded into the instrument, and 600 s (10 min) of scanning with continuous rotation was selected to ensure the highest achievable accuracy and precision. Each sample was tested in triplicate (or 9 replicates as needed, see below) by re-loading the composited and/or ground samples onto the pedestal independently for each run (and not simply re-reading the same preparation three or nine times). This proved to be valuable for capturing if/when sample heterogeneity might be high and therefore measurements might not necessarily be representative of the lot as a whole. Note that testing 9 replicates of 80 mg (i.e. 0.72 g) is comparable to the accepted practice of testing 0.5 g of spices in duplicate (i.e. 1.0 g) with EAM 4.7. Nevertheless, for both methods, it is advisable to increase the number of replicates if sample heterogeneity is encountered.

With this workflow established, we were able to process a sample (12 retail subsample composite) in about an hour from weighing to the final triplicate scan, with a cost of $2.48 USD per sample (Table 1). At this operating cost, once the instrument and common support equipment like grinders and balances are procured, the most expensive part of the protocol is the retail cost of the spice itself. Confident that the approach satisfied the need for a simple, rapid, and inexpensive workflow, we proceeded to evaluating its performance against FDA's guidelines for chemical screening methods.

3.2. Evaluating regulatory testing requirements

FDA has published detailed guidelines on criteria that are to be assessed for methods meant for compliance testing of regulated foods (Food and Drug Administration, 2019). For qualitative chemical screening methods, Section 2.3 outlines seven specific criteria that must be assessed: 1) minimum detectable concentration, 2) confirmation of identity, 3) sensitivity, 4) selectivity, 5) false positive rate, 6) false negative rate, and 7) ruggedness. We therefore set out to test our mHD-XRF screening method against each of these criteria in turn.

Minimum detectable concentration of the target is usually the first feature that needs to be established, because it alone determines whether the method is appropriate for screening purposes. Indeed, this is the reason why conventional polychromatic XRF is inappropriate for food screening. To establish a minimum detectable concentration, FDA guidance in Section 2.6 and Table 1 suggest a ‘spike’ level near the

Table 1
Practical Considerations for Selecting Lead Screening Methods.

Criteria	mHD-XRF	ICP-MS
Expense (test cost per sample, consumables only)	$2.48 USD	$65.81 USD[1]
Expense (cost of instrument)	~$60,000	>$250,000
Speed (time to result)	1 h 7 min	3 days
Sensitivity (lead LOD in spices)	0.044 ppm	0.010 ppm
Simplicity (operational steps)	weighing, grinding, reading (load and record result)	weighing, grinding, microwave-assisted acid digestion, centrifugation, gravimetric dilution, instrument calibration, data calculations, diligent acid-washing of quartz vessels
Simplicity (instrument requirements)	none (standard 120 V outlet, no control computer required, no support equipment, no ventilation needs, light and portable can be placed anywhere)	high-voltage/amperage specialty electrical outlet, acid-resistant ventilation, large weight-bearing bench needed, uninterrupted power supply, gas generator, chiller, vacuum pump, external gas tanks (helium and argon), dedicated control computer
Simplicity (service contract requirement and cost)	Not required but recommended ($3750/year)	Essential (~$30,000/year)

[1] Calculated at maximal capacity to reduce individual costs.

anticipated limit of detection, followed by an assessment of accuracy and precision.

However, XRF is a surface technique that measures lead in dry ground powders, which makes 'spiking' by impregnating a blank matrix with lead not feasible. Instead, we used a spice with a naturally occurring lead concentration of 0.058 ppm, which is very near the previously published spice LOD of 0.060 ppm (Johnson-Restrepo et al., 2025). Moreover, the current work is focused on spices, and when only one matrix is used, FDA guidelines require samples be analyzed in 9 replicates instead of 3 (Food and Drug Administration, 2019).

mHD-XRF analysis revealed a mean concentration of 0.056 ppm lead (Fig. 1), which is within 3.4% of the true value established by FDA EAM 4.7, and quite remarkable accuracy at such a low lead level. The standard deviation across 9 independent readings was 0.0145 ppm. This dataset resulted in a limit of detection (LOD), which is commonly defined in XRF studies as three times the standard deviation (Cifuentes et al., 2026; Johnson-Restrepo et al., 2025), of 0.044 ppm, or 44 ppb (Fig. 1).

Most screening techniques test for the presence of a contaminant (e.g. *Listeria monocytogenes*) whose mere presence results in a conclusion of adulteration. However, naturally occurring elements like lead are often present at some detectable level, but constitute adulteration only when present above a specific concentration. Therefore, in addition to establishing an LOD, accurate quantification of lead in spices is essential to this method being reliable for screening spices. To address this issue, we further evaluated the limit of quantification (LOQ), which is defined as 10 standard deviations, and found it to be 0.145 ppm, or 145 ppb (Fig. 1).

While it is common practice to treat spices as a single food type, spices are naturally diverse and thus could behave differently when tested for lead using mHD-XRF. In order to evaluate whether this common simplification is reasonable, we compared the LOD and LOQ calculated with 15 different spices with lead determined by EAM 4.7 to be below the mHD-XRF LOQ. Since under this assumption each spice would be considered a separate matrix, FDA guidelines require only 3 replicates. This analysis revealed the average LOD to be 0.064 ppm,

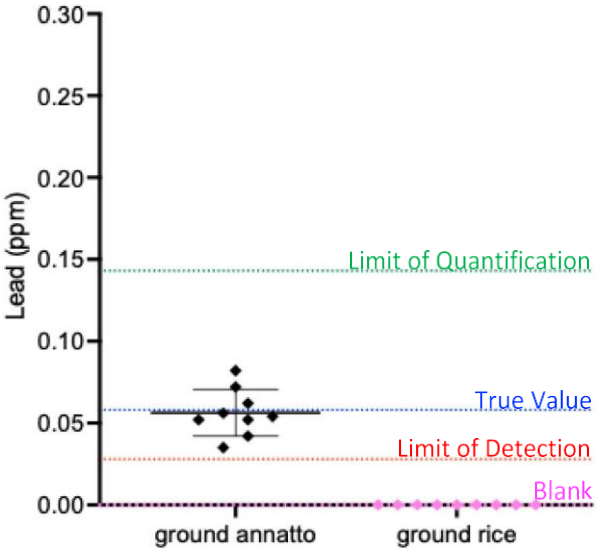


Fig. 1. Establishing limits of detection and quantification for lead in spices by mHD-XRF. A spice with lead near a previously established LOD (annatto, 0.058 ppm) was tested in 9 replicates by mHD-XRF. The mean and SD are shown in black, true value established by ICP-MS as a blue dashed line, LOD as a dashed red line, LOQ as a dashed green line, and true blank (ground rice) in pink. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

which is remarkably similar to 0.044 ppm as derived with just annatto (Fig. 2). The LOQ was also similar (0.212 ppm versus 0.145 ppm), thus verifying that, in general, considering spices a single matrix is acceptable for mHD-XRF testing.

While establishing an LOD and LOQ is necessary, it is not sufficient, because mHD-XRF is an inherently quantitative technique, yet we are developing mHD-XRF for qualitative screening. Since we would be using this screening method across a wide range of lead levels, we sought to evaluate the accuracy and precision across additional levels of lead. Again, since only one and not three matrices were being used, we satisfied the FDA guidelines by evaluating each sample in 9 replicates

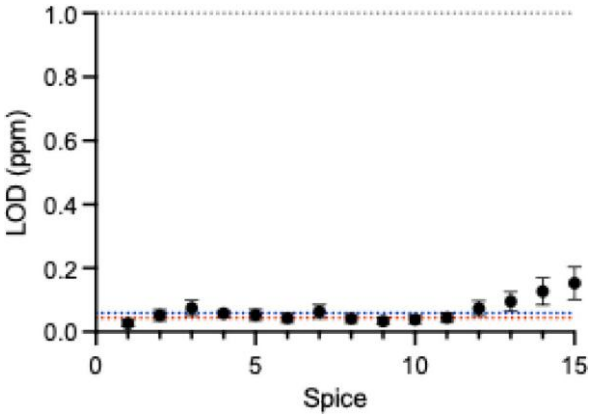


Fig. 2. Limits of detection compared across 15 spices. All spice samples that contained endogenous lead below the mHD-XRF LOQ of 0.145 ppm were used to calculate LODs for each individual spice. Plotted are the resulting LODs (+/— SD), with blue dashed line indicating the average LOD, red dashed line indicating LOD calculated with annatto alone (from Fig. 1), and grey dashed line indicating the target trigger level. Samples analyzed were: 1 = coriander, 2 = cinnamon, 3 = red chili, 4 = annatto, 5 = chili, 6,8,11–15 = paprika, 7 = Kashmiri chili, 9 = hot Hungarian paprika, 10 = turmeric. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(instead of 3 replicates with 3 different matrices). This analysis revealed excellent accuracy and precision of <20% at ~0.7 ppm, 1.2 ppm, and 2.5 ppm lead (Fig. 3), providing confidence that the accuracy and precision would be adequate across a range of lead levels in diverse spices.

Once we established the minimum detectable concentration, we moved on to the second requirement, which is confirmation of analyte identity. FDA recommends establishing chemical identity using mass spectrometry (Food and Drug Administration, 2019). Accordingly, we achieved this by either previously or concurrently analyzing all spices samples using ICP-MS with FDA EAM 4.7, which is an independent orthogonal approach that is approved by FDA as a confirmatory method. Specifically, the sample preparation procedures are entirely different (physical grinding of samples into fine powders versus complete digestion with acid, pressure, and heat for elemental analysis), as is the method of detection and quantification (X-ray fluorescence spectroscopy versus mass spectrometry). In all cases, lead was confirmed, with no outliers or false identifications.

3.3. Setting a threshold 'trigger level' for lead

The next set of 4 performance measures deal with true versus false positives and negatives. As such, we needed to set a 'trigger level' or 'threshold level' that would define a positive sample. The purpose of a screening method is to identify food products that warrant further testing for potential public health action. As such, we chose a 'trigger level' of 1.0 ppm, because a recent analysis of recalled ground cinnamon following the cinnamon applesauce outbreak (20, see 4. Discussion) revealed that the lowest concentration that led to a recall was 2.03 ppm (Food and Drug Administration, 2025a). As such, 1.0 ppm is sufficiently low to ensure products that would lead to public health action are captured, while being low enough to support aggressively identifying suspect products that could lead to action in the future as FDA progresses in its goal of achieving 'Closer to Zero' (Food and Drug Administration, 2025b).

In order to use a trigger level, we believed it important to establish uncertainty of this method such that samples that fall under the trigger level numerically but within the method uncertainty are not missed. To do so, we established uncertainty using a spice sample that had naturally occurring lead at near the trigger level, and found the uncertainty to be 0.06 ppm at a 99% confidence interval (Fig. 4). Since our trigger level thus proved to be 7 standard deviations from the LOQ, we were confident that false negatives would be exceedingly rare.

3.4. Key screening method performance measures

Sensitivity of a screening method is paramount to its usefulness, because for screening methods sensitivity does not refer to the minimal concentration of an analyte that can be detected, as it does for quantitative methods. Rather, sensitivity is defined across samples as the percentage of positive samples that are detected versus those that are missed in a sample dataset (Food and Drug Administration, 2019). This is calculated by dividing the number of true positives detected by the sum of the true positives plus false negatives (TP / [TP + FN]). With the retail set of 136 samples, no false negatives were encountered (Fig. 5), indicating that the sensitivity of the screening method is 100%.

In addition to sensitivity, selectivity is an important screening method performance parameter, because it measures the level at which the screening approach mistakenly includes false positives across samples. The parameter is defined at the true negatives divided by the sum of true negative plus false positives (TN / [TN + FP]). In this case, again, we encountered no false positives (Fig. 5), and as such the selectivity of the screening method was calculated to be 100%.

The most difficult metrics to establish for any screening method are the statistical false positive and false negative boundaries. The generally acceptable rate for a screening method to be useful is <5% false positive/negative rate at a 95% confidence level (Appendix 2B in FDA guidelines in 18). To establish this statistical boundary requires at least 59 samples above the trigger level to be analyzed to estimate the false negative rate. The challenge here is that, with a general violative rate of ~11% (Ishida et al., 2022), very few states are likely to have encountered enough positive retail samples to have the required sample dataset needed to evaluate this key parameter.

As such, we combined our positive spice samples ($n = 16$) collected over the past 5 years of testing with those from New York ($n = 47$) and Connecticut ($n = 12$). This provided a sample set of 75 spices with lead >1 ppm collected at retail in 3 states over a 13-year period and analyzed by ICP-MS. Analyzing this extensive sample set revealed no false negatives with a trigger level of 1.0 ± 0.06 ppm (Fig. 5), thereby establishing the false negative rate to be <5% at a 95% confidence level, as outlined

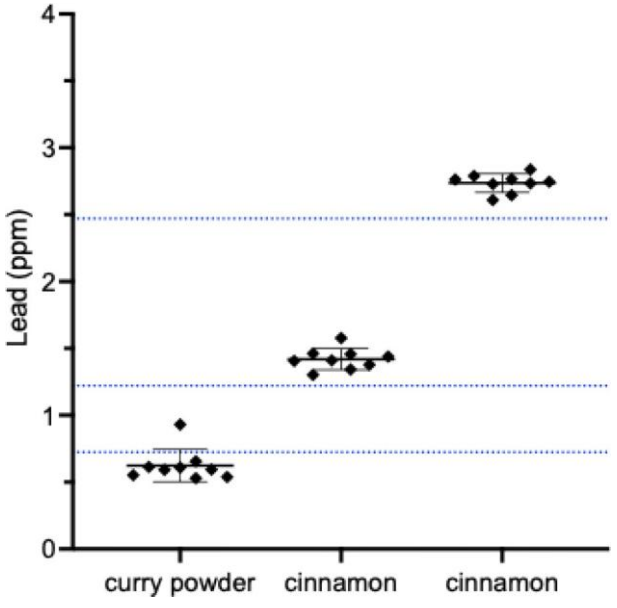


Fig. 3. Assessing accuracy and precision of lead quantification in spices by mHD-XRF. Spices with three different levels of endogenous lead were tested in 9 replicates by mHD-XRF. Mean and SD are shown, with true values established by ICP-MS shown as blue dashed lines. Note that although incomplete digestion can low-bias ICP-MS values, all values are nevertheless within 20% across the two platforms. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

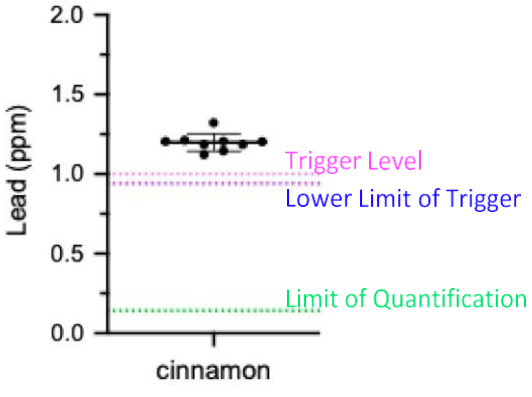


Fig. 4. Establishing trigger level uncertainty. A spice (cinnamon) with endogenous lead (shown as black points with mean +/— SD) near the chosen trigger level was assessed for uncertainty. LOQ is shown as a green dashed line, while the trigger level is displayed as a red dashed line, and a lower boundary as a 99% confidence interval in a purple dashed line. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

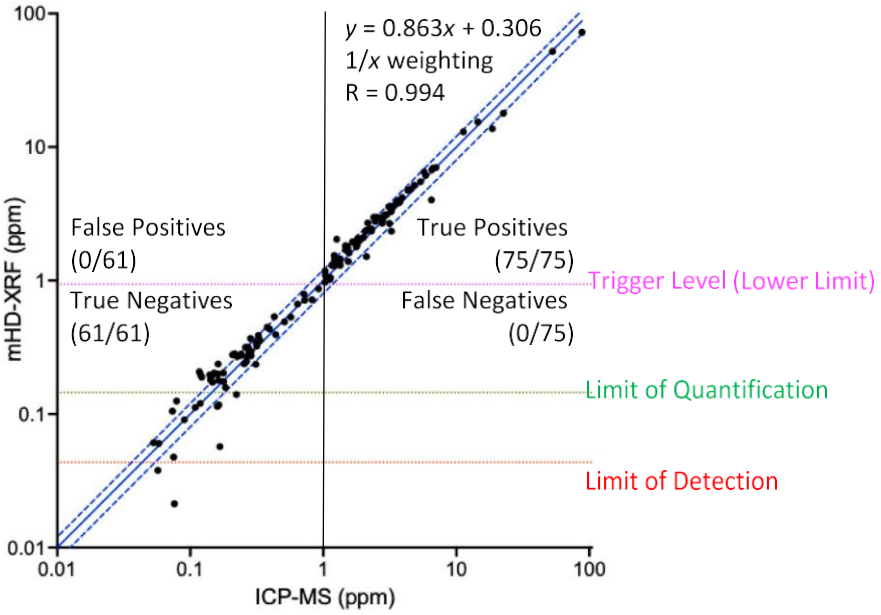


Fig. 5. Spice sample collection (*n* ¼ 136) assessed by ICP-MS versus mHD-XRF for correlation and true/false positives/negatives. Mean lead measurement values by ICP-MS and mHD-XRF are shown for the entire spice sample collection. The equation was fit with a linear model and 1/x weighting. A theoretical line of unity (slope of 1.0 and origin intercept) is shown in blue, with dashed lines indicating +/— 20% deviation. Note that both axes are logarithmic scale only to facilitate visualization of data spread. Lower 99% confidence interval of the trigger level (1.0–0.06 ppm) is indicated by a pink line, while the mHD-XRF LOQ and LOD are displayed as green and red dashed lines, respectively. True/false positive and negative quadrants (bound by pink and black lines) are labeled, and the number of samples falling into each is indicated. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

in Appendix 2B of FDA's guidelines (Food and Drug Administration, 2019). This is critical, because false negative samples would be excluded from any further analysis, and, as such, would be allowed to expose the public to harmful levels of lead.

False positives, although analyzed in a similar way, are not as critical, because these would be subjected to further ICP-MS analysis and thereby shown not to be a health concern. An elevated false positive rate would therefore waste resources by mandating further testing of samples that are actually not a health risk. Although this is certainly to be avoided, it must be considered in the current context of laboratories having no screening method, and thus having to subject *all* samples, sight unseen, to expensive and labor-intensive confirmatory testing. Therefore, even a terrible false positive rate of 50% would half wasteful laboratory testing. Nevertheless, we tested 61 spice samples below the trigger level and encountered no false positives, indicating that the true rate is also <5% at a confidence level of 95% (Fig. 5).

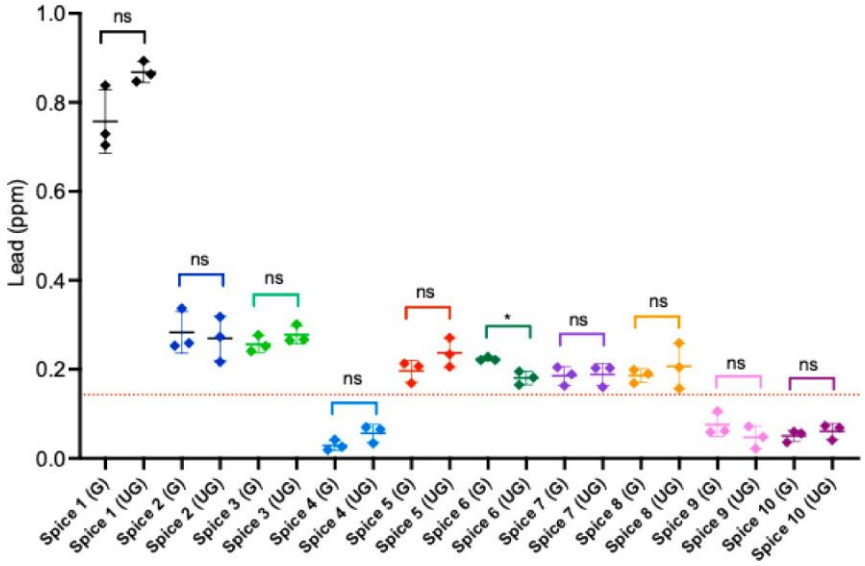


Fig. 6. Evaluating mHD-XRF method ruggedness. Ruggedness was assessed by examining 10 coarse spices before (UG, unground) and after (G, ground) homogenization. A paired student *t*-test revealed the following levels of statistical significance: Spice 1 (barberry): $p = 0.1248$, Spice 2 (paprika): $p = 0.7487$, Spice 3 (Cayenne pepper): $p = 0.4152$, Spice 4 (salad seasoning): $p = 0.1623$, Spice 5 (smoked paprika): $p = 0.1915$, Spice 6 (turmeric): $p = 0.0402$, Spice 7 (paprika): $p = 0.3949$, Spice 8 (paprika): $p = 0.4226$, Spice 9 (red chili): $p = 0.2941$, Spice 10 (chili): $p = 0.5893$. Dashed red line delineates the LOQ. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.5. Assessing method ruggedness

The final parameter that is to be assessed for screening methods is ruggedness, which is the ability of a method to withstand deliberate perturbation and nevertheless deliver meaningful results. While the MHD-XRF analyzer is a portable unit and could be tested in the field under a variety of conditions that could affect its ruggedness (e.g. electrical source stability, vibrations, heat, humidity), our intent was to use it in a meticulously controlled laboratory environment as a regulatory screening tool. As such, in our specific case we determined that a real-world perturbation is likely to be the degree of spice grinding. As such, we evaluated the effect of no grinding to full grinding on a spice sample set of all ten coarse spices that we had at in our sample set (Fig. 6). The median difference in ground versus unground was 16.1%, which was within acceptability criteria. Importantly, a paired student's *t*-test revealed no statistical significance between nine ground versus unground samples tested, while one experienced low significance ($p = 0.0402$). As such, we conclude the method is rugged to deliberate 'real-world' perturbations.

4. Discussion

Food safety systems in advanced countries like the United States are robust, such that generally 1% or fewer of foods present any health risk to the public. However, while this is a notable achievement, most people eat at least 3 meals comprised of multiple foods each day. At that level of consumption, even a 1% adulteration rate presents unacceptable risk (assuming 3 foods per meal x 3 meals per day x 365 days at a 1% risk constitutes ingestion of ~33 adulterated foods each year). Coupled with the particularly dangerous risk presented by lead due to exerting harmful and irreversible effects at low concentrations and bio-accumulating continuously, food safety systems must adopt a higher level of protection.

Screening methods have revolutionized food safety testing, because they are able to sift through many irrelevant samples that present no or acceptable health risks quickly and inexpensively to identify those rare ones that might not. In doing so they are able to reverse the 'victim of success' challenge posed by <1% adulterated rate, and facilitate meaningful rather than futile surveillance testing. While this has been a well-established approach in microbiology for decades, it is rare or virtually unheard-of in chemistry testing. We have sought to overcome this gap by establishing lead screening as a valid regulatory method for spices, which constitute a major vulnerability.

The recent outbreak of lead poisoning in children from cinnamon applesauce pouches has brought into sharp focus the need for such a screening capability for lead (Troeschel, Buser, Winquist, et al., 2025). This outbreak was detected in the worst way possible: elevated blood lead levels in children. The public health response was swift and organized, but laboratory testing could nevertheless have been improved.

Early on, state laboratories collected additional products and conducted lead testing to determine the scope of the outbreak. The Maryland public health laboratory was able to respond immediately, but even with overtime and weekend work, it took 4 business days to establish the levels of lead in ~20 retail products. In contrast, an established lead screening approach would have ruled-out safe products within hours. This would have saved time, dramatically increased testing capacity, and decreased costs by many-fold.

Even at the outset, the first state laboratory to test food samples could not run a specialized method like EAM 4.7, and therefore quantifying the actual levels of lead in the products had to wait until a laboratory competent in that method could be assigned. In this case, mHD-XRF even with a junior laboratorian would have yielded an estimate within an hour that would likely be ±20% of the EAM 4.7 value.

Thirdly, when the outbreak investigation turned to imported ground cinnamon, a lead screening method would have facilitated large-scale market surveillance at a fraction of the time and cost that was incurred. Even if 10% of imported ground cinnamon was adulterated and subject to public health action, the fact that all had to be tested using confirmatory methods like EAM 4.7 meant that testing cost was needlessly an order of magnitude higher than was required. Put another way, the same investment in resources could have yielded 10-fold more samples being tested and thus a much higher level of protection.

The cinnamon applesauce outbreak was ultimately caused by deliberate adulteration of imported cinnamon with lead chromate (Troeschel et al., 2025). While experienced analytical chemists are needed to investigate the root cause of these complex cases using sophisticated new methods (Kubachka, Witkowski, & Hurt, 2026), a strategic goal of our study was to have an Association of Public Health Laboratories (APHL)-CDC Public Health Laboratory Fellow (HB) without any laboratory experience perform the mHD-XRF screening described here. This proved to be very successful, indicating that good will and attention to detail, rather than years of sophisticated analytical chemistry training, are all that is needed for success.

Installing portable analyzers at multiple ports of entry could similarly enable import control staff to analyze samples themselves, which would create an opportunity for this analysis to become a requirement for product release into the nation's food supply. This has the potential to dramatically reduce the public's exposure to adulterated products, because they would be removed before they reach food store shelves, instead of relying on retail surveillance testing. Notably, while our studies provide strong reason to believe in the eventual success of this strategy, certain performance metrics should be reassessed in the field (e.g. ruggedness with field power supply, humidity, etc) under intended use conditions that could be different from our controlled laboratory environment.

We have deliberately focused on one metal, namely lead, in one matrix, namely spices, to develop and validate a mHD-XRF screening method so that it can be operationalized quickly for public good. But it should be noted that the approach taken here has proven to be robust, and warrants further analyte extension studies to other toxic elements, and matrix extension studies to other foods, that will likely lead to similar success.

Finally, while we have prioritized leveraging mHD-XRF to develop a screening method that would trigger testing of potentially adulterated samples by conventional methods, it has not escaped our notice that the strong performance of the mHD-XRF on a quantitative level indicates that it could eventually replace confirmatory testing by ICP-MS for such matrices as spices. Since a level-2 single laboratory validation study is insufficient to support such a dramatic change, we are therefore currently starting a level-4 multi-laboratory study comprised of an additional 10 laboratories. If successful, this extensive study would support adoption of mHD-XRF as a quantitative technique and potentially allow its incorporation as an AOAC International method.

An added benefit of this advance would not only be increasing surveillance scale and decreasing cost, but also freeing EAM 4.7 to focus on more vulnerable foods that lie below the LOQ of mHD-XRF, such as childhood foods that constitute the greatest public health vulnerability (Food and Drug Administration, 2025b).

CRediT authorship contribution statement

Hala Baidas: Writing – review & editing, Validation, Methodology, Formal analysis, Data curation. **Melikizat Begzad:** Writing – review & editing, Validation, Formal analysis, Data curation. **Sadia Muneem:** Validation, Supervision, Data curation. **Sinisa Urban:** Writing – original draft, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Hala Baidas, Sinisa Urban reports financial support was provided by Association of Public Health Laboratories. Sinisa Urban reports financial support was provided by US Food and Drug Administration. Sinisa Urban reports a relationship with Association of Public Health Laboratories that includes: funding grants and travel reimbursement. SU serves on the Association of Public Health Laboratories Human and Animal Food Committee, Environmental Laboratory Science Committee, 2026 Conference Organization Committee, and various thematic sub-workgroups. These are not paid appointments, but are national service on a professional association. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. The disclosures reported here are included because the prompts included disclosing professional relationships.

Acknowledgements

The authors are grateful to our colleagues Dr. Maria Ishida at the New York State Department of Agriculture and Markets, and Meghan Cahill at the Connecticut Agricultural Experiment Station, for generously providing their regulatory spice samples and associated ICP-MS testing data for this study. We are grateful to Lara Phillips for collecting spice samples in Maryland. The authors also wish to express our sincerest gratitude to Dr. Patrick Parsons of the New York Wadsworth Center for his lifetime of leadership in public health metals testing, and for introducing us to mHD-XRF. The authors are grateful to Dr. Zewu Chen, David Gibson, and Karen Horth at *Z*-Spec, Inc. for training and continuous support on the use of their mHD-XRF analyzers, and to Drs Kevin Kubachka, Patrick Grey, and Mike Farrow from FDA for reviewing and providing feedback on the single laboratory validation study. This work was partly supported by a fellowship from the Association of Public Health Laboratories (APHL) with funds awarded by CDC in NU60OE000104, and FDA grants U19FD007084 and U19FD008575 to SU. The Maryland Department of Health cannot endorse or recommend products, and as such any mention of specific products herein is simply information sharing for the purpose of scientific accuracy, and should not be construed as product recommendation or endorsement. The contents of this article are the sole responsibility of the authors, and do not necessarily reporesent the official views of MDH, APHL, FDA, CDC, or HHS.

Data availability

The datasets described in this article will be made available by request to the corresponding author after completion of the multi-laboratory study, and in accordance with departmental policies.

References

de Boer, D. K. G. (1989). Fundamental parameters for X-ray fluorescence analysis. *Spectrochimica Acta Part B: Atomic Spectroscopy, 44*, 1171–1190.

Centers for Disease Control and Prevention. (2025a). Childhood lead poisoning prevention. Recommended actions based on blood lead level. https://www.cdc.gov/lead-prevention/hcp/clinical-guidance/index.html.

Centers for Disease Control and Prevention. (2025b). Childhood lead poisoning prevention. About the data: blood lead surveillance. https://www.cdc.gov/lead-prevention/php/data/blood-lead-surveillance.html.

Chen, Z., & Gibson, W. M. (2002). Doubly curved crystal (DCC) X-ray optics and applications. *Powder Diffraction Journal, 17*, 99–103.

Chen, Z. W., Gibson, W. M., & Huang, H. (2008). High-definition X-ray fluorescence: Principles and techniques. *X-Ray Optics and Instrumentation, 2008*(10), Article 318171.

Chen, Z. W., & Wittry, D. B. (1998). Microanalysis by monochromatic microprobe x-ray fluorescence - physical basis, properties, and future prospects. *Journal of Applied Physics, 84*, 1064–1073.

Cifuentes, M. E. A., Rodriguez, Y., Avellaneda-Torres, L. M., Quiroga-Mateus, R., & Bravo, D. (2026). Assessment of monochromatic X-ray fluorescence spectrometry as a reliable analytical technique for cadmium quantification in cacao systems. *Spectrochimica Acta Part B: Atomic Spectroscopy, 235*, Article 107352. https://doi.org/10.1016/j.sab.2025.107352

Environmental Protection Agency. (2017). Proposed Modeling approaches for a health based benchmark for Lead in drinking water. https://www.epa.gov/sites/default/files/2017-01/documents/report_proposed_modeling_approaches_for_a_health_based_benchmark_for_lead_in_drinking_water_final_0.pdf.

Food and Drug Administration. (2019). Guidelines for the Validation of Chemical Methods for the FDA Foods Program, 3rd Edition. https://www.fda.gov/media/81810/download.

Food and Drug Administration. (2025a). More ground cinnamon products added to FDA public health alert due to presence of elevated levels of Lead. https://www.fda.gov/food/alerts-advisories-safety-information/more-ground-cinnamon-products-added-fda-public-health-alert-due-presence-elevated-levels-lead.

Food and Drug Administration. (2025b). Closer to zero: Reducing childhood exposure to contaminants from foods. https://www.fda.gov/food/environmental-contaminants-food/closer-zero-reducing-childhood-exposure-contaminants-foods.

Food and Drug Administration. (2026a). Bacteriological Analytical Manual. https://www.fda.gov/food/laboratory-methods-food/bacteriological-analytical-manual-bam.

Food and Drug Administration. (2026b). Compliance Program Guidance Manual 7304.019. https://www.fda.gov/media/142504/download.

Gray, P. J., Mindak, W. R., Cheng, J., & Kovalenko, A. (2025). Inductively coupled plasma-mass spectrometric determination of arsenic, cadmium, chromium, Lead, mercury, and other elements in food using microwave assisted digestion (version 1.3). . Food and Drug Administration Elemental Analysis Manual Chapter 4.7. https://www.fda.gov/media/191258/download?attachment

Ishida, M., Greene, V., King, T., Greenberg, J., Luker, J., Oglesby, D., Sheridan, R., & Trodden, J. (2022). Regulatory policies for heavy metals in spices – A New York approach. *Journal of Regulatory Science, 10*(1), 1–12. https://doi.org/10.21423/jrs-v10i1ishida

Johnson-Restrepo, B., Blain, E., Judd, C., Tysoe, A., & Parsons, P. J. (2025). New developments in monochromatic energy dispersive X-ray fluorescence instrumentation for monitoring toxic elements in food matrices: Advantages and limitations. *Radiation Physics and Chemistry, 234*, Article 112749. https://doi.org/10.1016/j.radphyschem.2025.112749

Kitov, B. I. (2000). Calculation features of the fundamental parameters in XRF. *X Ray Spectrometry., 29*, 285–290. https://doi.org/10.1002/1097-4539(200007/08)29:4<285::AID-XRS428>3.0.CO;2-C

Kubachka, K., Witkowski, M., & Hurt, D. (2026). Detection of lead (II) chromate in cinnamon. *Applied Spectroscopy Practica., 4*(1). https://doi.org/10.1177/27551857261425269

Marguí, E., Queralt, I., & de Almeida, E. (2022). X-ray fluorescence spectrometry for environmental analysis: Basic principles, instrumentation, applications and recent trends. *Chemosphere, 303*, Article 135006.

Neltner, T. (2017). Children's lead exposure: relative contributions of various sources. EDF Health. Retrieved from https://blogs.edf.org/health/2017/12/15/childrens-lead-exposure/ Accessed December 15, 2017.

Revenko, A. G., & Pashkova, G. V. (2023). X-ray fluorescence spectrometry: Current status and prospects of development. *Journal of Analytical Chemistry, 78*, 1452–1468.

Shahid, M., Dumat, C., Khalid, S., Schreck, E., Xiong, T., & Niazi, N. K. (2017). Foliar heavy metal uptake, toxicity and detoxification in plants: A comparison of foliar and root metal uptake. *Journal of Hazard Materials, 325*, 36–58.

Troeschel, A. N., Buser, M. C., Winquist, A., et al. (2025). Investigation of lead and chromium exposure after consumption of contaminated cinnamon-containing applesauce — United States, November 2023–April 2024. *MMWR. Morbidity and Mortality Weekly Report, 74*, 239–244. https://doi.org/10.15585/mmwr.mm7414a2