Mitigating aflatoxin exposure to improve child growth in Eastern Kenya (MAICE)
Statistical analysis plan

Version July 28 2016

1. Introduction

Mitigating aflatoxin exposure to improve child growth in Eastern Kenya (MAICE) is a cluster-randomized controlled trial with the goal to determine whether consumption of aflatoxin has a direct causal impact on child linear growth during the period from before birth to age 24 months. Below we describe the MAICE statistical analysis plan whilst data collection is still ongoing. Data collection is expected to be completed by end of 2016.

2. Overview of study

2.1. The MAICE study

The MAICE study is a cluster-randomized controlled trial that compares an intervention designed to reduce aflatoxin exposure compared to a control group.

The intervention consists of two components: swap and stock. In the swapping component, households are visited monthly by trained staff of Caritas Meru, a nongovernmental organization that works with farmers in the study area, and offered rapid aflatoxin testing of any stored maize that the household plans to consume over the next 2 months. If the household agrees, a composite sample of at least 150 grams of flour or 300 grams of whole kernels is taken from six regions of the container in which the identified maize is stored, and ground using a manual grinder, so that at least 70 % of the sample passes through a 20-mesh sieve, in accordance with US Grain Inspection, Packers and Stockyards Administration (GIPSA) aflatoxin testing protocols. A rapid aflatoxin test is then conducted using a 10-gram sub-sample of the homogenized composite sample using the GIPSA-verified Romer (Romer Labs®, Inc., USA) AgraStrip rapid test with a 10 parts per billion (ppb) detection threshold, according to manufacturer instructions, as described in Appendix 1.

Any maize found to contain over 10 ppb aflatoxin (the Kenyan regulatory limit for aflatoxin contamination) is replaced with an equal amount of maize that has been tested and found to contain less than 10 ppb aflatoxin ("safe maize"). This component of the intervention has been in place since July 2013.

The stockist component of the intervention was introduced after trial commencement, in January–February 2014, in response to the observation that many households were accessing the majority of the maize they consumed through the market. This was unexpected and arose due to an unusually poor maize harvest in the study area in 2013. Household maize purchases are typically small and frequent, so the swapping intervention was missing a large portion of maize consumed. In the stockiest intervention, Caritas Meru supplies maize containing less than 10 ppb aflatoxin to at least one shopkeeper in each of the intervention villages. In geographically larger villages, safe maize is supplied to multiple shopkeepers to ensure this maize is accessible to all study participants. Participating households are encouraged to purchase this tested maize, which is offered at the lowest price of maize currently for sale in the village. This encouragement is through an initial village meeting, as well as the monthly swapping visits. Participants have been provided with a laminated ID card displaying their name and unique identifying number. To ensure that an adequate supply of safe maize is available to study participants, stockists are asked to sell maize only to those presenting a study ID card, and to record all sales in a tracking form which includes a field for the household unique identifying number. This enables sales information to be linked with household outcomes to estimate the reduction in aflatoxin exposure at the household level. Participants in villages assigned to the intervention group were consuming untested, purchased maize for up to 12 months before the stockist component of the intervention was introduced, half of the study duration of 24 months. Due to the risk that these participants were exposed to the full intervention for an insufficient length of time to have an impact on the primary outcomes of interest, an additional wave of recruitment was added to replace this wave of recruits. Wave-1 households in the intervention group

continued to receive the intervention for a 2-year period, according to the original protocol, but this group will be excluded from follow-up data collection.

2.2. Study population

The study population consists of pregnant women after the fifth month of gestation and their offspring. Within each of the study villages, enrollment into the study was conducted in 6 waves, each 4 months apart. In each wave, women in the fifth to final month of pregnancy (by the woman's estimate) were invited to participate in the study[1]. All offspring of the pregnant women became part of the study. In case of twins or triplets, the children's names were ranked alphabetically and the first child selected as the study child[2].

2.3. Inclusion criteria

2.3.1. Inclusion criteria (pregnant women)

Pregnant women were eligible for inclusion in the study if all of the following criteria were met:
a) At least 18 years of age at the time of enrollment;
b) Fifth to final month of pregnancy (self-report);
c) Willing to sign an informed consent;

2.3.2. Inclusion criteria (infants)

Infants were eligible for the study if they were delivered from the pregnancy for which their mothers were enrolled into the study. In case the study child died or moved outside of the study area, and a child between 12 to 35 months was living in the household, the child was included in the study. As there are very few of these replacement children, they will be dropped from the analyses.

2.4. Exclusion criteria

2.4.1. Exclusion criteria (pregnant women)

Pregnant women were excluded from the study if they were not residing in the study site (i.e. one of the 56 villages in Meru and Tharaka-Nithi Counties in Kenya) or were not willing to participate in the study surveys.

2.4.2. Exclusion criteria (offspring)

In case of twins or triplets, children whose names were not first alphabetically were excluded from the study.

2.5. Objectives

The primary objective of this study is to determine whether consumption of aflatoxin has a direct causal impact on child linear growth during the period from before birth to age 24 months. We also aimed to explore which social, economic, demographic, health, nutritional characteristics may modify the effects of the intervention. These factors could modify the exposure to the intervention or modify

[1] We informed village leaders that we were looking for women who were in their 5th or later month of pregnancy. Pregnancy was then verbally verified with respondents, who were also asked which month of pregnancy they were in. Reported gestational age was recorded. Each of these respondents were then asked to identify other pregnant women in the village.

[2] Including all children from twins and triplets would require controlling for potential intra-household correlation, which complicates the analyses without adding much power.

the (biological) effect of intervention exposure on the outcomes of interest.

The null hypotheses that this study aimed to reject were that:

- The two groups of children (i.e. intervention and control) will not differ significantly in mean length at endline[3].
- The two groups of children (i.e. intervention and control) will not differ significantly in serum aflatoxin albumin (AF-alb) level at endline (as determined using HPLC analysis according to methods detailed in Appendix 2).

2.6. Outcomes

2.6.1. Primary outcomes

The study has two primary outcomes, namely:

- Linear growth status of the child (delivered by enrolled woman) assessed using LAZ[4] at endline (both as a continuous variable and as the prevalence of stunting)

Child serum AF-alb level at endline (log-transformed continuous variable).

2.6.2. Secondary outcomes

Secondary outcomes for the <u>child</u> (delivered by the enrolled woman) include:

- Additional indicator of linear growth at endline: Length-for-age difference[5]

- Indicators of health status:
  - Prevalence of any reported symptoms[6] of illness in the past two weeks at endline;
- Indicators of child development
  - Age-standardized score on assessments of motor, language, and cognitive skills
- Outcomes using midterm data:[7]
  - Linear growth status of the child (delivered by enrolled woman) assessed using LAZ at midterm
  - Length-for-age difference
  - Child serum aflatoxin AF-alb level at midterm;

Pathways through which the intervention may have affected growth (especially through consumption of high risk foods and non-impacts on value of consumption) will be assessed through analysis of the following <u>household</u> level variables:

- Consumption of maize procured through the study and non-study sources based on data from the swapping questionnaire
- Total household food consumption at endline (energy per day adjusted for

[3] The endline survey is conducted approximately 24 months after the enrollment survey. Children will thus be approximately 20 to 24 months.

[4] The WHO 2006 Child Growth Standards will be used for age-and-sex standardization of the child weight, length, and weight-for-length, and BMI outcomes.

[5] Length-for-age difference is the difference between the child's length and the median length for the child's age and sex from the WHO 2006 International Growth Standard. See Leroy, J. L., Ruel, M., Habicht, J.-P., & Frongillo, E. A. (2014). Linear Growth Deficit Continues to Accumulate beyond the First 1000 Days in Low- and Middle-Income Countries: Global Evidence from 51 National Surveys. The Journal of Nutrition, 144(9), 1460–6. This outcome will be used in any longitudinal analysis as it is statistically incorrect to use HAZ for such analysis.

[6] We do not (initially) plan to look at individual illness (e.g. fever, diarrhea, etc)

[7] Study households enrolled in waves four to six participated in a mid-term survey conducted at the time of the endline survey in the third enrollment wave.

household composition);
- Household food and non-food expenditure at endline.

2.7. Safety outcomes

We assess safety by evaluating adverse events and increases in serum AF-alb as explained in the following sections.

2.7.1. Adverse events

Data on adverse events are collected through routine monitoring and through the scheduled household surveys. The routine monitoring consists of any reports on adverse events reported by study field staff or by households through the contact information provided on the informed consent form. Adverse events that will be monitored through this system (as requested by the trial's Data Safety Monitoring Board) include deaths, hospitalization, and any illness, injury, or complaints.
The household survey data will be used to assess childhood illness, anthropometry, death, and the serum samples collected during the surveys to determined serum AFB-alb. It must be noted that there is a lag time of several months between survey data collection[8] and the availability of the survey data for tabulation[9].

2.7.2. Tabulation of adverse events

Adverse events and results on main outcomes will be tabulated and reported to the DSMB. If adverse events are likely related to the intervention, the TSC and DSMB will be informed immediately to discuss trial continuation. The tabulated information on the main outcomes will also be used by the DSMB to decide if the motivating study question has been answered.

2.8. General principles underlying the analyses

- The primary analysis will be intention-to-treat, i.e. outcomes for all children enrolled will be analyzed according to the group to which they were assigned. Within this approach, three sets of analysis will be conducted:
    1) All enrollment waves combined;
    2) In light of the strengthening of the intervention after trial commencement (i.e. the inclusion of the stockist intervention, see section 2.1.), we will estimate the impact separately for children who were exposed to the improved intervention later in life (waves 2 and 3), and those were exposed early in life (waves 4, 5, and 6).
    3) (Gestational) age at onset of the full intervention[10] (swap and stock)– this analysis is similar to the wave-based analysis described above, but in addition to changes in the protocol also exploits individual-level variation in its implementation;
- In addition to the primary analyses, two separate dose-response analyses will be conducted within the intervention group. These will make use of variation over time in the

[8] At both midterm and endline, mothers of children who looked ill or weak were advised to take their child to the nearest health facility. Transportation was facilitated by either taking them to the health facility of providing money for the bus fare.
[9] After completion of a survey round, it takes approximately 3 months to ship and analyze the serum samples.
[10] As the full intervention (i.e. the addition of the stockist component) was rolled out after the trial had started, the age at which children started to be exposed to the full intervention varied. Some children were exposed starting in utero and others started being exposed after they were born. Gestational age will be inferred from actual birth dates recorded on children's government-issued health cards assuming a 40-week gestation (if the health card is not available, child's birth date as reported by the mother will be used).

implementation of the intervention, background aflatoxin levels, and, in the second approach, variation in household-level compliance. Statistical power for these analysis will depend on the extent of within-cluster variation in gestational age at onset and compliance with the intervention. The sources of variation used in each of these approaches are as follows:

4) Estimated reduction in aflatoxin exposure over time based on differences in implementation and environmental factors outside the control of study households. This approach uses the average reduction in aflatoxin exposure for each child as the treatment variable, which will be calculated using the child's (gestational) age at onset, and the month- and village- specific average reduction in aflatoxin exposure. To calculate the reduction in aflatoxin exposure, we will use data on the proportion of total maize consumed among the treatment group that was obtained through the study each month and average month and village-level estimated AF levels in household stores based on data from the swapping intervention;
5) Estimated reduction in aflatoxin exposure over time based on differences in household compliance, study implementation and environmental factors. In addition to the variation utilized under approach 4), the household-specific proportion of obtained through the study at the household level will be used to construct the treatment variable. This final approach, while subject to selection bias, will provide additional supportive evidence for the ITT findings.

- All tests on the effect of the treatment will be one-sided, at the 5% level of significance.
- Since this is a cluster-randomized trial, all statistical analyses will account for within-cluster correlation of outcomes.
- We will report the number of observations used in each analysis.

3. Design

3.1. Sample size and power

To determine sample size, the minimum detectable effect (MDE) for exposure to treatment for the full 24-month period was set to 0.3 length-for-age Z-score (LAZ), based on the magnitude of effect sizes that have been obtained with known effective nutrition interventions. This was then adjusted for partial exposure to the intervention of participants recruited in the first year of the study (prior to introduction of the stockist component) to give an MDE of 0.281. Sample size calculations were performed to achieve an alpha of 0.05 and power of 80 % to detect a difference of means in LAZ between children in intervention and control villages using a 1-sided test. Based on anthropometric data collected in similar conditions, the standard deviation of the LAZ of children was assumed to be 1.28.

The study attempted to enroll all children born in study villages during the enrollment period. Birth rates from 2008 to 2009 Kenya Demographic and Health Survey and populations of the study villages from the 2009 Kenya Census were used to estimate the average number of births per village between enrollment waves spaced 4 months apart.

Each village was defined as a cluster, including all households recruited from that village across all waves. For the purpose of the power calculations, the number of households per cluster per recruitment wave was assumed to be equal across villages, at 3.3 births per study wave. This was based on mean village size and pregnancy rates observed in the 2008-2009 Kenya DHS data for the Eastern Region. In reality, due to significant variation across villages in population size, the total number of households enrolled per cluster ranged from 2 to 100 (mean=22, median=16) over the course of five waves.

In our power calculations, we first used Stata's sampsi command to calculate the total un-clustered sample size required to power the study. Next, we assumed an intra-cluster correlation (ICC) for child growth z-scores of .05. We identified a logistically feasible range for the number of recruitment waves and repeated the process described below for each assumed value for number of waves.

We calculated a design effect as 1 + (N-1)*ICC, where N=total number of households per cluster (3.3* the assumed number of waves). We adjusted the required sample size by multiplying it by the ratio of the design effect to the proportion of the sample assumed to be successfully tracked for follow-up (.85). The required number of clusters for a given number of recruitment waves was calculated by dividing the adjusted required sample size by the number of waves per cluster.

Assuming an intra-cluster correlation for LAZ of 0.05, attrition of 9 %, and that 15 % of variation in the outcome would be explained by baseline variables, the estimated total number of participants required to achieve the study objectives was 924 across 56 villages.

3.2. Informed consent

Informed consent was obtained from all participants as follows. The consent form was read to each potential participant by an enumerator at the initial visit. The participant was then asked to sign the document if she agreed to participate. All pregnant women who gave informed consent in this way were enrolled in the study.

3.3. Treatment groups

A random number generated assigned a rank to each village enrolled in the study. Villages randomly assigned to the 28 highest ranks were assigned to the treatment arm.

3.4. Data collection and follow-up

Data collection at baseline and follow-up occurs at participants' homes through face-to-face interviews using handheld tablets on which an electronic data collection form is programmed. A pre-coded survey was administered to the expectant mother immediately after enrollment, her height and weight were measured, and self-reported month of pregnancy was recorded. Expectant mothers were also asked to provide a venous blood sample to be analyzed for serum aflatoxin. A similar survey is repeated during endline data collection at 24 months after enrollment. A mid-term survey was conducted among participants enrolled in the fourth through sixth waves at the time of the endline survey in the third enrollment wave. At each follow-up visit, the length and weight of the child in utero at baseline (reference child) is recorded, and a venous blood sample is taken from the child for serum AF-alb analysis.

3.5. Interim data analysis

The impact of the primary outcomes will be assessed when all endline data are available. . However the Data and Safety Monitoring Board (DSMB) and Institutional Review Board (IRB) can request interim analysis of child anthropometric outcomes and serum AF-alb level by study group. In case of any evidence that the intervention is leading to adverse outcomes, or to an overwhelmingly statistically significant treatment effect prior to the planned study end date, the DSMB will inform the Trial Steering Committee and advise on the appropriate course of action.

3.6. Definitions

Term	Unit of measurement	Severe	Moderate	Mild	General
Stunting	Length-for-age z-score (WHO 2006)	< -3	>= -3 to < -2	>= -2 to < -1	< -2

4. Statistical analysis

4.1. Study flowchart

A participant flow diagram will be prepared in accordance with the CONSORT 2010 guidelines. In line with the guidelines, we will report for each study arm:

- The numbers of clusters who were randomly assigned (and number of participants in these clusters)
- The number of clusters that received intended treatment (and number of participants in these clusters)
- The number of clusters that were analyzed for the primary outcome (and number of participants in these clusters)

4.2. Procedures for data cleaning

Data cleaning is performed at various points during data collection:

- The electronic data collection form used in the field includes range and consistency checks, automatic skip patterns, and pop-up messages if questions are left blank.
- Stata syntax was written by project staff to identify improbable and missing values; data problems, to the extent possible, were resolved by checking with the field team or by a repeat home visit whenever possible.

4.3. Outliers

We will identify outliers by visually inspecting Box plots and/or histograms of individual continuous variables, and scatterplots of related variables. Two possible courses of action will be followed:

- Clearly impossible or implausible values for age, height, and survey responses will be corrected if possible, or recoded to missing if correction is not possible.
- Given the typically skewed of the AF-alb variable and the statistical and aflatoxin expertise of the TSC, decisions on outliers for this variable will be discussed with the TSC for input.
- Plausible or possible outliers will be kept. Variables with outliers might be transformed, and in an extreme situation (for example if the number of outliers exceeds 5% of the number of observations), a sensitivity analysis will be done to assess to what extent these outliers influence the results.

4.4. Software

All analyses will be done using Stata (version 14 or higher, StataCorp, TX, USA).

4.5. Background characteristics of participants and baseline comparisons

For all variables measured, the available values at enrollment, i.e. before the start of the intervention, will be considered as baseline characteristics. The background characteristics of subjects who completed the study will be presented in Table 1, by treatment group. However, for specific papers, only selected background characteristics considered relevant to those papers will be presented.

Analysis of background characteristics will be conducted as follows:

- Categorical variables: frequencies and percentages, as appropriate. Percentages will be calculated based on the number of participants for whom data are available.
- Continuous variables: mean and SD or median, IQR and range, as appropriate.
- Where data for certain participants are missing, the number of participants included in the analysis will be indicated.

Balance across treatment groups will be assessed by regressing each variable on randomized treatment allocation, clustering standard errors at the village level using the Huber-White approach.

4.6. Potential effect modifiers

The following variables will be considered as potential effect modifiers:

- Proxy variable for household socio-economic status or wealth[11]
- Enrollment wave
- (Gestational) age at onset of improved intervention

Non-categorical effect modifiers will be transformed into categorical variables by classifying observations. The specific effect modifiers that will be considered in each analysis are provided in the sections below.

4.7. Timing of measurement of outcome variables

All outcomes are measured approximately 24 months after study enrollment (i.e. when the child is approximately 20 to 24 months of age[12]). In addition, study households enrolled in waves four to six participated in a mid-term survey conducted at the time of the endline survey in the third enrollment wave. These mid-term surveys took place respectively 19, 15 and 11 months after study enrollment for waves four, five and six, respectively.

4.8. Analysis of the effect of the intervention

The analysis of the effect of the intervention will begin with testing the null hypothesis of no difference among the two study arms on primary and secondary outcomes using linear regression. We will conduct crude analysis (not controlling for any covariates), analyses controlling for pre-specified (see section 4.8.3 below) covariates, and analyses of "reduced" models, i.e., models retaining only those covariates that are statistically significant.

The effects of potential effect modifiers will be assessed with an interaction term in the linear regression model. Interactions with significant p-values and/or that may be substantively important will be further examined with stratified analyses, estimation of separate regression lines, or estimation

[11] An asset index will be developed using a simple count of the number of assets owned, a simple count of the different types of assets owned, and principle component analyses of the number of assets owned. Each variable will be dichotomized and the variable explaining the largest amount of variation in the child linear growth outcome at endline will be used in the analyses.

[12] Length was measured in children under 24 months of age and standing height in children older than 24 mo of age, in line with the WHO recommendation. Children outside of the 20-24 month age range during endline data collection will be included in the primary analysis.

of adjusted means at key points of the covariate, in order to understand the nature of the effect modification.

4.8.1. Primary outcomes

The linear regression model to determine the effects of the intervention on length and serum aflatoxin AF-alb level will be fitted using Stata's regress command. Results will be summarized in tables, providing the mean and SD by study group, the estimated difference in means (and the 95% confidence bound) between the study groups. AF-alb level may be logarithmically transformed prior to analysis due to the skewness that typically characterizes the distribution of this variable.

4.8.2. Secondary outcomes

The procedure described in section 4.8.1 will be followed for secondary outcomes. The first analyses will focus on the outcome at endline. For observations with repeated measures (i.e., household in waves 4 through 6 in which a midterm survey was conducted), may also employ linear mixed-effects models (Stata mixed command) if the model's assumptions are satisfied.

4.8.3. Selection of covariates and effect modifiers in the analysis of primary outcomes

Below we list the two primary outcomes to be analyzed (both at endline), and indicate the covariates and effect modifiers that will be used in the analysis. Covariates will be included if they were unbalanced at baseline, explain variability in the outcomes, or both. Final analyses will include "reduced" models that retain only covariates that are statistically significant. Each of the effect modifiers will be considered in a separate regression model to avoid collinearity.

Primary outcome variable	**Analysis**	**Covariates to be considered**	**Effect modifiers**
LAZ and stunting	Linear regression	- Household level: • Proxy variable for household food insecurity (using FANTA's HFIAS) • Proxy variable for household diet quality (using household dietary diversity) • Proxy variable for household socio-economic status or wealth (see above) - Individual level: • Child sex • Child age • Maternal height • Maternal education • Maternal age • Household head education • Number of adult equivalents in household - Other: • Enrollment wave	*See 4.6*

Primary outcome variable	Analysis	Covariates to be considered	Effect modifiers
		•Season of birth	
Child serum AF-alb	Linear regression	- Household level: •Proxy variable for household food insecurity (using FANTA's HFIAS) •Proxy variable for household diet quality (using household dietary diversity) •Proxy variable for household socio-economic status or wealth (see above) - Individual level: •Child sex •Child age •Maternal education •Maternal age •Household head education •Number of adult equivalents in household - Other: •Enrollment wave •Season	*See 4.6* *(enrollment wave will not be considered as an effect modifier for this outcome, nor will (gestational) age at onset of the improved intervention)*

4.8.4. Selection of covariates and effect modifiers in the analysis of secondary outcomes

A similar analytic approach will be followed as for the primary outcome variables. Covariates will be included if they were unbalanced at baseline, explain variability in the outcomes, or both. Final analyses will include "reduced" models that retain only covariates that are statistically significant. Each of the effect modifiers will be considered in a separate regression model to avoid collinearity.

Secondary outcome variable	**Analysis**	**Covariates to be considered**	**Effect modifiers**
See 2.6.2.	Linear regression	- Household level: • Proxy variable for household food insecurity (using FANTA's HFIAS) • Proxy variable for household diet quality (using household dietary diversity) • Proxy variable for household socio-economic status or wealth (see above) - Individual level: • Child sex • Child age • Maternal height • Maternal education • Maternal age • Household head education • Number of adult equivalents in household - Other: • Enrollment wave • Season	*See 4.6*

4.9. Data transformation

Continuous outcomes will be assessed for normality and will be transformed when necessary. If normality and homogeneity of variances cannot be optimized through a suitable transformation, analysis will be done on ranked data.

4.10. Confidence intervals

In line with the one-sided tests, estimates of treatment effects will be accompanied by a 95 % confidence bound where possible.

Appendix 1: Maize aflatoxin test procedures

1. 10 g of the ground, homogenized sample is weighed into a clean jar that can be tightly sealed.
2. 20 ml of 70% methanol, 30% water extraction solution is added and the jar is sealed.
3. This mixture is vigorously shaken for 1 minute.
4. The sample is allowed to settle
5. The test micro well is placed in a micro well-holder
6. Using a single channel pipette, 50ul of assay diluent is transferred into the micro well. The coating conjugate in the micro well is then dissolved by pipetting the content up and down 5 times.
7. 50 µl of sample extract is added the micro well, mixing the content in the well by pipetting it up and down 3 times.
8. A test strip is placed into the well for 5 minutes.
9. Results are interpreted immediately.
10. If there is no line in control zone, the test is invalid and the sample should be re-tested by using a valid test strip.

Appendix 2: Analytical methods for measurement of aflatoxin B albumin

Serum samples are analyzed with HPLC-fluorescence method (Qian et al., 2010, 2013), including measurement of albumin and total protein concentrations for each sample, digestion with protease to release amino acids, concentration and purification of aflatoxin-lysine adduct, and separation and quantitation with HPLC system. Specifically, thawed human serum samples are inactivated for possible infectious agents via heating at 56 ℃ for 30 minutes, followed by measurement for albumin and total protein concentrations using procedures modified as previously described. A portion of each serum sample (150 µL) or recorded volume are digested by pronase (pronase: total protein, 1:4, w: w) at 37°C for 3 h to release AFB-Lys. AFB-Lys in digests are further extracted and purified by passing through a Waters MAX SPE cartridge, which is preprimed with methanol and equilibrated with water. The loaded cartridge is sequentially washed by water, 70% methanol, and 1% ammonium hydroxide in methanol at a flow rate of 1ml/min. Purified AFB-Lys is eluted with 2% formic acid in methanol. The eluent is vacuum-dried with a Labconco Centrivap concentrator (Kansas City, MO) and reconstituted for HPLC-fluorescence detection.

The analysis of serum AFB-Lys adduct is conducted in an Agilent 1200 HPLC-fluorescence system (Santa Clara, CA). The mobile phases consists of buffer A (20 mM NH4H2PO4, pH 7.2) and buffer B (100% Methanol). The Zorbax Eclipse XDB-C18 reverse phase column (5 micron, 4.6 x 250 mm) equipped with a guard column is used. Column temperature is maintained at 25°C during analysis, and a volume of 100 µL is injected at a flow rate of 1 mL/min. A gradient is generated to separate the AFB-Lys adduct within 25 min of injection. Adduct is detected by fluorescence at maximum excitation and emission wavelengths of 405 nm and 470 nm, respectively. Calibration curves of authentic standard are generated weekly, and the standard AFB-Lys is eluted at approximately 13.0 min. Quality assurance and quality control procedures include simultaneous analysis of one authentic standard in every 10 samples and two quality control samples daily. The limit of detection is 0.4 pg/mg albumin.