A Nonlinear Mixed Modeling Method to Analyze Assay Data and the Effects of Exposures of Two Indoor AeroAllergens During Infancy on Children at Age Three: The CCAAPS Cohort
A Nonlinear Mixed Modeling Method to Analyze Allergen Assay Data and the Effects of Exposures Two Indoor Aeroallergens During Infancy on Children at Age Three: The CCAAPS Cohort

A dissertation submitted to the
Division of Research and Advanced Studies
of the University of Cincinnati
in partial fulfillment of the
requirements for the degree of
DOCTORATE OF PHILOSOPHY (Ph.D.)
in the Division of Epidemiology and Biostatistics
of the Department of Environmental Health
of the College of Medicine
2009
by
Juan Liang
MA in Economics, University of Oregon, 2004
BA in Economics, University of Colorado, 2002
Abstract

In chapter II, a nonlinear mixed modeling method is used to analyze immunoassay data. First, a nonlinear mixed four parameter logistic model is used to estimate the concentration-response relationship in the standard samples. Second, sample concentrations are calibrated by using the estimated relationships from standard samples. This nonlinear mixed modeling method is applied on allergens sample data in the Cincinnati Childhood Allergy and Air Pollution Study (CCAAPS). It provides wider detection ranges than a log-log linear modeling method. More samples with low/high concentrations are estimated. In chapter III, the estimated home cat and dust mite allergen concentrations from samples which were collected at the first year of child’s age are used to explore the exposure-response relationships of two home aeroallergens at age one and three clinical outcomes at age three. Higher levels of dust mite allergen concentration are found to be a risk factor to develop persistent wheezing at age three and to be protective to develop SPT positivity to cat allergen at age three. African American children and non-African American children are observed having disproportionally high prevalence rates of persistent wheezing and SPT positivity at age three and also several demographic characteristics. From public health perspective, physicians and parents should be aware of these findings in order to improve asthma management and patient care.
Acknowledgements

I would like to gratefully and sincerely thank my advisors, Dr. Linda Levin and Dr. Paul Succop for their expertise and guidance. I am also indebted to Dr. Grace LeMasters and Dr. David Bernstein, who allowed me to use the data from the CCAAPS study.

I am particularly thankful to my parents, my husband and my daughter Lisa. All these accomplishments become more meaningful with their love, support and inspiration.
Appendix II. Testing Homogeneity of Intra-plate Variance ...................................................... 51
Appendix III. Lower Limit of Detection (LLOD) and Upper Limit of Detection (ULOD) Calculation.......................................................................................................................... 52
Appendix IV. Assessing Experimental Conditions which Affect Dog Assay Variability .. 53
Appendix V. Comparison of Different Logistic Models................................................................. 58
Appendix VI. Statistical Software.................................................................................................. 61
Appendix VII. R \textit{nlme} Random Effect Covariance Modeling ............................................. 65
Appendix VIII. Results Comparison: Results from NLME Model vs. Results from the Log-log Model................................................................................................................................. 66
Appendix IX. Comparison of Parameter Estimates from the NLME Model.............................. 68
Appendix X. Comparison of Several Standard Curve Fitting Approaches............................. 70
Appendix XI. Cat Allergen Assay and Laboratory Dilutions ...................................................... 75
Bibliography.................................................................................................................................. 77
Chapter I. Introduction

1. Statistical Methods of Assay Data Analysis

During the past 20 years, bioassay technology has achieved major advances. Efficiency in laboratories has been improved tremendously. More sophisticated and systematic statistical methodologies are demanded to analyze and interpret laboratory results. Enzyme-linked immunosorbent assay (ELISA) is an analytical method that is widely used by laboratories to quantify concentrations of substances in various biological and environmental samples. Substance concentrations in the samples are obtained through two steps. The first step is estimation of a standard curve or dose-response curve by using laboratory generated data. The second step is to use the inverse of the estimated standard curve to predict concentrations of the samples. Besides quantifying substance concentrations in the samples, assay performance is also very critical. Assay performance can be characterized by its reliability and sensitivity. Reliability of an assay is measured by assay precision profiles. Sensitivity is measured by limits of detection (LOD).

Picture1. Description of Assay Analysis
2. Cincinnati Childhood Allergy and Air Pollution Study and Home Dust Sample Collection

The Cincinnati Childhood Allergy and Air Pollution Study (CCAAPS) is an ongoing prospective birth cohort study of infants born to an atopic parent. The parent was considered atopic if he or she had at least one allergy or asthma symptom and had a positive SPT to one of 15 aeroallergens. The study includes families residing in the Greater Cincinnati Metropolitan Region of Southwestern Ohio and Northern Kentucky. Infants were identified from birth certificates and their parents’ SPT results determined whether they were eligible to enter into the study. Children underwent a physician health examination and skin prick testing to a panel of 15 aeroallergens at yearly clinic visits. In addition, at each visit a parent is administrated a standardized questionnaire which inquired respiratory symptoms including wheezing during the past year and other in depth information.

Dust samples were collected from flooring materials in the primary activity room using a vacuum cleaner. Antigens were extracted from the dust sample and analyzed using a commercial antibody-based enzyme-linked immunosorbent assay (ELISA) (Indoor Biotechnologies, Inc., Charlottesville, VA). Cat (Fel d 1), house dust mite (Der f 1), and Cockroach (Bla g 1) allergens were analyzed by capture assay using monoclonal antibodies. Dog (Can f 1) and Alternaria antigens were analyzed using polyclonal antibodies.
3. Exposure-Response Relationships of Indoor Aeroallergens and Allergic Health Outcomes

Accurate and précised estimated allergen concentrations from superior statistical methods can be used to assess the exposure-response relationships of allergen exposures and allergic health outcomes. Aeroallergen sensitization and asthma related airway symptoms among young children have been studied in various populations from different geographic locations to explore exposure-response relationships between indoor home aeroallergens during early infancy and various allergic immune responses in childhood. The exposure-response relationships were varied which probably related to the child’s age, the allergen concentration ranges of the living environments and the types of the allergic health outcomes.
Chapter II Statistical Methods of Assay Data Analysis

1. Objective

This chapter describes a non-linear heterogeneous variance mixed effects modeling approach for analyzing ELISA assay data. The calibration of concentrations of substance levels in the samples and the characterization of assay performance are described in detail.

The non-linear heterogeneous variance mixed effects modeling methods are applied to allergen data obtained from indoor dust samples, provided by the Cincinnati Childhood Allergy and Air Pollution Study (CCAAPS). Home dust samples are analyzed by ELISA to obtain concentrations of dust mite (Der f1), dog (Can f1), cat (Fel d1) and roach (Blag1) allergens. Examples of these methods are applied to the determination of dust mite allergen concentrations.

2. Background

Standard Curve Estimation

The first step is estimation of standard curves or dose-response curves by using laboratory generated data. There are many different approaches to estimating a dose-response relationship in bioassays. The estimation procedures in this paper assume a monotonic relationship between the dose and the response, i.e. the local slope of the standard curve does not change sign, and there is no local minimum or maximum. In addition, the response associated with the dose is continuous, although the response in bioassay can be quantal or binary. Standard curve estimation methods for quantal response assays can be found in Govindarajulu (2001) [1] and Finney (1978) [2].
The simplest standard curve model is linear regression. Sometimes the dose and/or the response are transformed prior to the analysis. Linear regression has the advantage of graphical and arithmetic simplicity. Whatever functional relationship is assumed, the linear portion of the curve is most important. The assumption of linearity, however, has been found to hold only for the middle portion of a laboratory standard curve data. The non-linear portions at the lower and upper asymptotes are ignored or simply truncated.

A polynomial approach has also been applied. Hayward and coworkers (1991) [3] used third order polynomial regression to model the relationship between the response and log dose. They showed the $R^2$ from the polynomial was improved compared to the linear model. However, polynomials are not widely practiced in bioassay due to their instability at the asymptotes (Lee Dipaolo Ji, 2000) [4] and mathematical difficulty of inverting the curves. A nonparametric approach to standard curve estimation has also been applied, but has found to be non-informative compared to parametric approaches.

Finney (1976) [5] recommended the use of a family of (nonlinear) sigmoid equations, i.e. a four parameter logistic (FPL) model, which was flexible enough to describe many assay systems. Furthermore, he proposed various variance functions to describe the unequal variability of the response over the entire dose range. This paper will employ Finney’s heterogeneous FPL model and will use a mixed modeling technique to combine multiple plates in the same assay to improve the precision of variance parameter estimates from the FPL.

*Description of the ELISA Assay*

---

1 To get the unknown concentration from the standard curve, the polynomials require to be solved. For third order polynomial there are three roots, two of which are physically insignificant. (Hayward, et. al. 1991)
Many standard curve estimation procedures initially appeared in a radioligand assay background. ELISA assays have almost identical characteristics except that the response in ELISA assays is monotonically increasing instead of decreasing. Each ELISA assay contains several microtiter plates, which are analyzed on the same day and assumed to be handled under the same experimental conditions, e.g. reagent, temperature and humidity. For each plate there is a laboratory “standard kit” which contains concentrations obtained by sequentially diluting purified solutions of the substance of interest. Optical densities (ODs) are detectable signals measured by the laboratory, which are related to the standard concentrations. In practice, most laboratories report background ODs where no substance is present. In the context of the dose-response curve, substance concentration is the dose, and OD is the response. More than one microtiter plate is usually analyzed in each assay. Each plate is assumed to have the same concentrations. In each plate the ODs are measured repeatedly on a continuous scale over multiple serial dilutions of concentrations.

3 Statistical Methods

The primary goal of bioassay analysis is to estimate substance concentrations in environmental samples. Substance concentrations in environmental samples are obtained through two steps. The first step is estimation of the standard curve or dose-response curve by using laboratory generated data. The second step is to use the inverse of the estimated standard curve to calculate substance concentrations in environmental samples.

The relationship between the logarithm of concentration and the OD response may be described by an “S” shaped curve called the four parameter logistic (FPL) model. The FPL model is the following:
\[ y_j = \mu_j + e_j = f(x_j, \beta) + e_j; \quad f(x_j, \beta) = \beta_1 + \frac{(\beta_2 - \beta_1)}{1 + \exp(\beta_1 (\log x_j - \beta_1))}, \]

where \( y \) is the OD response, \( \mu \) is the mean OD response, \( x \) is the concentration, \( e \) is the error term and \( j \) indexes the concentration level. The error term \( e_j \) associated with the response \( y_j \) is assumed to have a normal distribution with a zero mean and a variance which is proportional to the mean OD response.

The FPL model is not only a very flexible model for fitting ELISA assay data, but also the parameters in the FPL model are each interpretable. The OD at an infinite concentration or at saturation is represented by \( \beta_1 \). The OD at zero concentration is represented by \( \beta_2 \). The logarithm of the concentration which corresponds to the 50th percentile of the OD (i.e., the log EC50) is represented by \( \beta_3 \). So, if \( x \) approaches zero, \( y \) approaches \( \beta_2 \). If \( x \) approaches infinity, \( y \) approaches \( \beta_1 \). If \( \log(x) = \beta_3 \), \( y = (\beta_1 + \beta_2)/2 \). \( \beta_4 \) is the slope at the midpoint between \( \beta_1 \) and \( \beta_2 \) for the linear portion of the curve; the absolute value of \( \beta_4 \) should be close to one, because a one to one relation is expected in the linear portion of the standard curve.

For most assays, the variation of the OD responses within each plate (intra-plate variation) is related to the average OD level. Therefore, instead of naively assuming homoskedasticity or constant variance (\( \text{Var}(y_j) = \text{Var}(e_j) = \sigma^2 \)), a variance function is employed to describe the heterogeneous intra-plate variation. This improves the precision of the regression parameter estimates, as well as the accuracy of precision profiles and limits of detection (LOD). Reciprocals of estimated variances will be used as weighting factors to estimate the parameters (\( \beta_1 \), \( \beta_4 \)) of the FPL model. The variance function \( \text{Var}(y_j) \) is assumed to be described by a function which is proportional to a power of the
mean response, \( \mu_j \), as recommended by Finney (1976 and 1977) [5, 6]. Thus the variance is a function of the regression parameters (\( \beta_1 - \beta_4 \)) through \( \mu_j \), \( \theta \) is the power parameter and \( \sigma \) is the scale parameter.

\[
\text{Var}(y_j) = \sigma^2 g^2(\mu_j, \theta), \quad g = \mu_j^\theta, \quad \mu_j = \text{E}(y_j) = f(x_j, \beta).
\]  

(2)

The estimation process is described in Appendix I.

**Pooling Information across Plates within the Same Assay Using a Non-linear Mixed Effects Model (NLME)**

The determination of assay reliability and sensitivity depend on variance parameter estimates. These may be poorly estimated when they are based on information from a single plate due to small sample size (Davidian, 1993) [7]. To improve variance estimates, different plates within the same assay are often combined using a mixed effects model (NLME) when variance estimates across plates in the same assay can be assumed to be equal. The assumption of homogeneity of variance can be tested through two Chi-square statistics (Zeng, 1997) [8]. These two chi-square tests are described in the Appendix II.

The heterogeneous variance nonlinear mixed model (NLME) is

\[
y_{ij} = f \left( x_{ij}, \beta_i \right) + e_{ij}, \quad \text{Var}(y_{ij}) = \text{Var}(e_{ij}) = \sigma^2 g^2(x_{ij}, \beta_i, \theta), \quad e_{ij} \sim N(0, R_i(\beta_i, \theta))
\]

(3)

where \( y_{ij} \) is the OD response, \( x_{ij} \) is the concentration, \( i \) indexes the plate, \( j \) indexes the concentration level, \( f(.) \) is the FPL model, and \( e_{ij} \) is random error for plate \( i \) and concentration \( j \). In the mixed model, \( \beta_i \) is a 4x1 vector containing the FPL parameters, \( \beta_{i1}, \beta_{2i}, \beta_{3i}, \) and \( \beta_{4i} \). \( \beta_i = \beta + b_i \), where \( \beta \) is a vector of fixed effects, and \( b_i \) is a random effect.
with $E(b_i) = 0$, $Cov(b_i) = D_i$, $b_i \sim N(0, D_i)$. The random effect, $b_i$, is independent of $e_{ij}$.

In this paper $R_i$ is defined as a positive definite diagonal matrix, i.e. $cov(e_{ij}, e_{i'j'}) = 0$, for all $i \neq i'$ and $j \neq j'$; a common intra-plate variance-covariance structure is found.

Parameter estimation of the NLME model may be carried out in R and SAS; it is required to specify starting values for the unknown regression parameters. It is very important that starting values are close to true parameter values. The initial starting values of the $\beta$s should be based on plots of the data, given the interpretation of the parameters.

*Calibration of Concentrations in Environmental Samples*

The primary purpose of the above analyses is to estimate substance concentrations in environmental samples. Parameters from the FPL standard curve analyses are used to estimate these concentrations. ODs of each environmental sample are measured by the laboratory at usually two to four dilutions of the original environmental sample. The dilutions allow ODs to be measured at several points along the standard curve. The goal is to get at least one OD value in the linear portion of the curve where concentrations are more accurate and precise. Recall the meaning of the parameters of the FPL model: $\beta_1$ is the OD at infinity concentration or at saturation. $\beta_2$ is the OD at zero concentration. $\beta_3$ is the logarithm of the concentration which corresponds to the 50th percentile of the OD (i.e., the log EC50) and $\beta_4$ is the slope at the midpoint between $\beta_1$ and $\beta_2$ or the linear portion of the curve.

From each observed OD, an estimated concentration may be obtained by inverting the fitted FPL standard curve. The formula for the estimated concentration $\hat{x}$ and its variance $\text{var}(\hat{x})$ are
\[ \hat{x} = h(y, \hat{\beta}) = \exp\left[ \hat{\beta}_i + \log\{ (\hat{\beta}_i - y) h(y - \hat{\beta}_i) \} / \hat{\beta}_i \right] \]
\[ \text{var}(\hat{x}) = h_i^2 (y, \hat{\beta}) \hat{\sigma}^2 (y, \hat{\theta}) + h'_\beta (y, \hat{\beta}) \hat{V}(\hat{\beta}) h_\beta (y, \hat{\beta}) \] 

(4)

where \( \hat{V}(\hat{\beta}) \) is the covariance matrix of \( \hat{\beta}_i \) from the NLME results, and \( i \) indexes the plate number.

The predicted concentration for a particular sample is based on the weighted average of estimated concentrations from all estimable dilutions of the same sample. The distribution of sample concentrations is skewed to the right and assumed to follow a log normal distribution, therefore a weighted geometric mean approximation to the sample concentration is employed, \[ \log \hat{x} = \sum w_k \{ \log(\text{dilution}_k * \hat{x}_k) / \sum w_k \} \], where \( k \) indexes the dilution level (usually \( k = 1 - 4 \)), and \( w_k \) is the weight factor for dilution \( k \), where
\[ w_k = \frac{1}{\text{var}(\log \hat{x}_k)} \]

and \( \text{var}(\log \hat{x}_k) \) is \( \text{var}(\hat{x}_k) \) in (4), with
\[ h(y, \beta) = \beta_i + \log\{ (\beta_i - y) / (y - \beta_i) \} / \beta_i \]. Thus, the dilution with the smaller variance is given more weight and a larger variance is given a smaller weight. The corresponding confidence interval for each estimated sample concentration \( \hat{x} \) is calculated as
\[ \exp(\log \hat{x} \pm Z_{\alpha/2} \text{var}(\log \hat{x}) \) where \[ \text{var}(\log \hat{x}) = \sum w_k^2 * \text{var}(\log \hat{x}_k) / (\sum w_k)^2 \]. This confidence interval is asymmetric, which takes into account the skewed distribution of \( \hat{x} \) and is more appropriate than other symmetric intervals (Belanger, 1996) [9].

Based on the calibration function from \( h(y, \beta) \), any \( y \) or OD outside \([\beta_2, \beta_1]\) will give a negative \( (\beta_2 - y) / (y - \beta_1) \). In this paper, to avoid negative values for \( (\hat{\beta}_2 - y) / (y - \hat{\beta}_1) \), the dilution(s) with OD outside \( \hat{\beta}_2, \hat{\beta}_1 \) for each sample are excluded. If the ODs in all the dilutions from a same sample are outside \( \hat{\beta}_2, \hat{\beta}_1 \), that sample is labeled as not estimable.
The concentration of that sample can be replaced by its (lower limits of detection) LLOD if the ODs in all dilutions are too small and by its upper limits of detection (ULOD) if the ODs in all dilutions are too large.

Assay Performance Evaluation Using Precision Profiles and Detection Limits

The precision profile is a summary of assay variability over the entire range of standard concentrations and is a measure of the reliability of the assay. It contains coefficients of variation (CV) of concentrations of standard curve data which can be used to identify concentrations of biological samples that are less reliable. Usually a CV is considered acceptable if it is less than 20%. Precision profiles are calculated by using the formula for the estimated concentration $\hat{x}$ and its variance $\text{var}(\hat{x})$ in (4). In this paper we calculate the precision profiles of standard curves from the variances of individual plates and pooled variances across plates for each assay.

Assay detection limits describe assay sensitivity. Sample concentrations above and below the limits of detection cause loss of sensitivity. This can pose significant challenges in epidemiology studies due to the possible reduction in the efficiency of effect estimation and loss of statistical power in hypothesis testing.

The lower limit of detection (LLOD) is defined as the lowest concentration, at which the OD can be distinguished from the OD of a zero concentration. It is a threshold above which the substance can be reported to be present in the sample (Diamandis, 1996) [10]. A common and convenient way to find the LLOD is to use the lowest concentration which is three standard deviations above the zero concentration. However, based on the definition of the LLOD, it is straightforward to use a one-sided t test to find the LLOD. This is the method used in this paper and is described in the Appendix III.
After obtaining the estimated LLODs, samples with extremely low substance levels will be given substance levels at either the LLOD or the median LLOD. Therefore, data that has been set to a missing value due to sample substance levels below the LLOD will be imputed. The number of environmental samples for which concentrations can be estimated is therefore increased, with resulting increase in power for hypothesis testing in epidemiologic studies.

There is no universal agreement on the definition of ULOD. Gottschalk (2005) [11] defined ULOD as the highest concentration which can be distinguished from an infinite concentration. To avoid concentrations above ULOD, laboratories may extend the standard curve or dilute environmental samples.

4. Application of Assay Methodology to a Dust Mite Assay

The methods are applied to dust mite allergen data obtained from indoor dust samples, provided by the Cincinnati Childhood Allergy and Air Pollution Study (CCAAPS). Table 2.1 shows the standard curve data from one plate of a dust mite assay, and Figure 2.1 is the graphical display of these data. This dust mite assay has four plates, each with a concentration range from 0.49 to 250 μg/ml. For each plate, at each concentration level, two optical densities are observed. The data of the whole assay and its figure are provided in the supplemental table and figure.
Table 2.1. One Plate from a Dust Mite Assay Standard Curve Data with Ten Standard Concentration Levels

<table>
<thead>
<tr>
<th>Plate # Standard Concentration (ng/ml)</th>
<th>Plate 1</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OD1</td>
<td>OD2</td>
<td></td>
</tr>
<tr>
<td>250.00</td>
<td>2.38</td>
<td>2.60</td>
<td></td>
</tr>
<tr>
<td>125.00</td>
<td>2.36</td>
<td>2.43</td>
<td></td>
</tr>
<tr>
<td>62.50</td>
<td>2.29</td>
<td>2.25</td>
<td></td>
</tr>
<tr>
<td>31.25</td>
<td>1.91</td>
<td>1.93</td>
<td></td>
</tr>
<tr>
<td>15.63</td>
<td>1.30</td>
<td>1.34</td>
<td></td>
</tr>
<tr>
<td>7.81</td>
<td>0.79</td>
<td>0.84</td>
<td></td>
</tr>
<tr>
<td>3.91</td>
<td>0.38</td>
<td>0.42</td>
<td></td>
</tr>
<tr>
<td>1.95</td>
<td>0.20</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>0.98</td>
<td>0.10</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>0.49</td>
<td>0.05</td>
<td>0.05</td>
<td></td>
</tr>
</tbody>
</table>
Figure 2.1. Graphical Display of Standard Curve Data from Plate 1 in the Dust Mite Assay in Table 1.
Results of fitting separate non-linear FPL models to each plate’s data are shown in Table 2.2. The $\beta$s are the regression parameter estimates; $\theta$ is the power of the mean in the variance function, and $\sigma$ is the scale parameter of the variance function. Coefficients of variation (CV) are calculated as the ratio of the standard deviation of plate estimates divided by the average of plate estimates for each parameter. Differences in the magnitude of each regression parameter and variance parameter estimate across plates are reflected in the coefficients of variation (CV) in the last row of Table 2.2. $\beta_2$, which is the estimated OD response at the lower asymptote, has the largest variability. The relatively large magnitude of the CV for $\sigma$ is primarily due to having small sample sizes (only ten concentration levels for each plate).

Table 2.2. FPL Parameter Estimates (One Plate At a Time) of a Dust Mite Assay

<table>
<thead>
<tr>
<th>Assay Plate</th>
<th>$\beta_1$ Estimate</th>
<th>$\beta_2$ Estimate</th>
<th>$\beta_3$ Estimate</th>
<th>$\beta_4$ Estimate</th>
<th>$\theta$ Estimate</th>
<th>$\sigma$ Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plate 1</td>
<td>2.59</td>
<td>0.02</td>
<td>2.7</td>
<td>1.27</td>
<td>0.79</td>
<td>0.04</td>
</tr>
<tr>
<td>Plate 2</td>
<td>2.42</td>
<td>0.03</td>
<td>2.56</td>
<td>1.31</td>
<td>0.59</td>
<td>0.07</td>
</tr>
<tr>
<td>Plate 3</td>
<td>2.55</td>
<td>0.03</td>
<td>2.7</td>
<td>1.26</td>
<td>1.02</td>
<td>0.05</td>
</tr>
<tr>
<td>Plate 4</td>
<td>2.53</td>
<td>0.00</td>
<td>2.77</td>
<td>1.15</td>
<td>1.04</td>
<td>0.06</td>
</tr>
<tr>
<td>Average</td>
<td>2.52</td>
<td>0.02</td>
<td>2.68</td>
<td>1.25</td>
<td>0.86</td>
<td>0.05</td>
</tr>
<tr>
<td>Variance</td>
<td>0.08</td>
<td>0.01</td>
<td>0.09</td>
<td>0.07</td>
<td>0.21</td>
<td>0.02</td>
</tr>
<tr>
<td>CV</td>
<td>0.11</td>
<td>5.00</td>
<td>0.11</td>
<td>0.21</td>
<td>0.53</td>
<td>2.83</td>
</tr>
</tbody>
</table>

Precision profiles usually are displayed graphically by plotting CVs versus predicted concentrations. Figure 2.2 shows the precision profiles of the dust mite assay. For all plates the coefficients of variation (which are proportional to the standard deviation) increase as the concentrations increase, confirming the appropriateness of the log transformation for variance stabilization. Table 2.3 provides concentration ranges where the CV criterion for the reliability of predicted concentration is satisfied. Plate one has
the best precision (greatest interval) and plate two has the worst precision (smallest interval). For plates one and two, estimated concentrations outside (0.45, 84.02) and (0.77, 42.61), respectively, are not reliable. Due to small sample sizes, precision profiles have limited accuracy when based on data from individual plates.
Figure 2.2. Precision Profiles of the Four Plates for the Dust Mite Assay. The criterion of reliability is a CV less than or equal to 0.2.
To improve the precision of variance parameter estimates, a non-linear FPL mixed effects model (NLME) may be applied by combining the data from all plates when estimates of sigma and theta are approximately equal. In this example, the assumption of homogeneity of variance parameters is valid, as determined by the chi-square tests described in Appendix II, and a NLME analysis is performed, which provided more stable estimates of the regression and variance parameters of the FPL model fit to the data from all plates.

Table 2.4 Fixed Effect Parameter Estimates from NLME

<table>
<thead>
<tr>
<th>Assay Plate</th>
<th>$\beta_1$ Estimate</th>
<th>$\beta_2$ Estimate</th>
<th>$\beta_3$ Estimate</th>
<th>$\beta_4$ Estimate</th>
<th>$\theta$ Estimate</th>
<th>$\sigma$ Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pooling 4 Plates</td>
<td>2.52</td>
<td>0.02</td>
<td>2.68</td>
<td>1.25</td>
<td>1.00</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Table 2.4 shows the fixed effect parameter estimates from NLME. Compared to the estimates in Table 2.2, estimates of the four fixed effects parameters, $\beta$, in Table 2.4 are exactly the same as the average estimates from the individual plates. However, the estimates of the variance parameters are modified by pooling the four plates. The estimate of $\sigma$ is 0.03 versus 0.04 to 0.07 in Table 2. The estimate of $\theta$ is 1.00, which is in the range 0.59 to 1.04 in Table 2.

The estimated random effect covariance matrix of the $\beta$s in the mixed effects model is
The diagonal elements are the estimated variances of the regression parameters $\beta_1$ through $\beta_4$. The CVs of $\beta_1$ through $\beta_4$ are 0.03, 1.58, 0 and 0.08. The between plate relative variability (CV=1.58) is largest at the lower asymptote ($\beta_2$). This is smaller than the CV of the same parameter from the analysis of the individual plates in Table 1 (CV=5.0).

The precision profile obtained from NLME provides 0.77 to 89.11 µg/ml as estimates for CVs ≤ 0.2. Compared to Table 3, the overall intra-assay precision using pooled variance parameter estimates has better (wider) coverage. This indicates that predicted concentrations are more precise and more reliable using pooled information across plates within the same assay, compared to only using the individual plate data.

Precision profiles are used to assess the reliability of an assay, from which the concentration range with contained variability can be identified. From the precision profile of this dust mite assay, any sample concentration above 89.11 µg/ml is subject to a huge error. The validity of any further hypothesis testing will be affected by errors in estimating sample concentrations. The statistical significance of the hypothesis test also can be reduced.

The lower (LLOD) and upper (ULOD) limits of detection are calculated from the pooled data, 1.52 µg/ml and 202.10 µg/ml, respectively. Compared to the concentration range from the precision profile in Table 4, the ULOD is not reliable due to a lack of precision,
since it is the above the higher limit of 89.11 μg/ml. The LLOD, however, is within the range and therefore is considered reliable.

In Appendix VIII, the results of the non-linear analyses are compared to those obtained from linear regression analysis, a common method for determining assay concentrations. The precision of estimated sample concentrations and assay sensitivities are improved, and more concentrations are estimable by using the current methodology, i.e. more samples are within the range of the lower and upper limits of detection.

5. Discussion

This paper describes and implements a combination of two methodologies for analyzing allergen assay standard data. The first method estimates regression parameters of a four-parameter logistic curve separately for each plate in the assay using non-linear regression analysis. The second method employs a non linear mixed effects model and combines the data from all plates in order to estimate a common variance of the data, thus increasing the precision of the predicted concentrations of environmental samples.

A four parameter logistic (FPL) model is known to fit assay data when the relationship between the concentrations and the optical densities is sigmoidal in shape. The validity of the FPL is usually not checked by the laboratory prior to the application to a specific assay. A limitation of this paper is that adequate selection of the FPL model is not addressed prior to applying the methods to the estimation of the concentrations of allergen data. The identification of the best model in nonlinear regression analyses was reported to be inconsistent by Bunke and coworkers (1999) [12]. These authors developed several different goodness of fit tests for this purpose, and suggested that the selection of the best model should not only be based on goodness of fit tests alone, but on
the intended use of the model, including the accuracy of concentration calibrations and estimation of the lower and upper limits of detection of assay-type data. In this paper several different models are compared to the heterogeneous FPL model (Appendix VIII) based on two criteria: maximization of $R^2$, a commonly used goodness of fit statistic, and precision of estimation of the parameters determining the lower limit of detection of the assay. The models to which the heterogeneous FPL model is compared are linear and cubic polynomial models, the three parameter logistic model, and the homogeneous four-parameter logistic model. Data from a dust mite allergen assay are used for these comparisons. The $R^2$s are obtained when each model is fit to the data from the four plates of the assay. For two assay plates, the three-parameter logistic model has the largest $R^2$ values. The heterogeneous FPL model has the highest $R^2$ for the other two plates. The heterogeneous FPL model, however, provides lower asymptotes for each plate. Based on a combination of these results, the heterogeneous FPL model is considered adequate for the methodology that is applied to these dust mite allergen data. An appropriate goodness of fit test to compare parametric nonlinear models can be developed in future research.

The programs used were developed in the Statistical Analysis System (SAS) and the R package and the applied programs are discussed in Appendix VI, which includes a comparison of the software programs used in this paper to several other software packages and programs.
Bibliography


Chapter III Health Outcome Analysis Using Home Aeroallergen Concentrations during Infancy

1. Introduction

Aeroallergen sensitization and asthma related airway symptoms among young children have been studied in various populations from different geographic locations to explore exposure-response relationships between indoor home aeroallergens during early infancy and various allergic immune responses in childhood. Early exposure to home pet allergens was shown to be protective against having positive IgE, positive SPT, wheezing and asthma when exposures were low [1, 2], risk for positive IgE when exposures were high [3], or have no effect on wheezing and sensitizations [4, 5]. The effects of early exposure to home dust mite allergen were different among studies as well. Some showed it had no clear association with positive IgE, wheezing and asthma [3, 4, 6], some showed a positive effects on asthma and wheezing [7], and some showed dust mite allergen and skin sensitization had bell-shaped relationships [8].

In terms of statistical methodology, most previous studies either assumed simple linear relationships between allergens and clinical outcomes or categorized the allergen levels to investigate nonlinear relationships. There are drawbacks of both methods. Modeling home allergens as a simple linear variable omitted the nonlinear exposure-response relationship. However, categorizing the allergen levels and assuming within each category all homes were homogeneous ignored a possible trend within each category. Categorizing the allergen levels also caused difficulty to detect any statistical significant nonlinear dose-response. Usually to show a nonlinear exposure-response relationship, the allergen levels often had to be divided to many subsets, such as using quintile or even
deciles. However, without a sufficient number of subjects, this could lower the power to detect any statistical significant results.

This study included children born to atopic parents who were enrolled in the Cincinnati Childhood Allergy and Air Pollution Study (CCAAPS) cohort. The CCAAPS study was designed to investigate the environmental factors that may be placing children at increased risk of developing allergies and/or asthma. In previous studies of this cohort, gender, race, number of siblings, exposure to cigarette smoke, parental asthma condition, and visible mold at home have been found to relate to sensitizations, wheezing and asthma [9, 10, 11].

The primary objective of this study is to apply a flexible statistical modeling approach to investigate potential non-linear exposure-response relationships between two indoor allergens (cat and dust mite) during infancy and allergic clinical outcomes including persistent wheezing and allergen specific SPT positivity at age three in this CCAAPS cohort. A linear spline technique is employed to explore the potential nonlinear relationships between home allergens and clinical outcomes and quantify trends within and between high and low allergen levels.

2. Methods

Study Cohort

The Cincinnati Childhood Allergy and Air Pollution Study (CCAAPS) is an ongoing prospective birth cohort study of infants born to an atopic parent. The parent was considered atopic if he or she had at least one allergy or asthma symptom and had a positive SPT to one of 15 aeroallergens. The study includes families residing in the
Greater Cincinnati Metropolitan Region of Southwestern Ohio and Northern Kentucky. Infants were identified from birth certificates and their parents’ SPT results determined whether they were eligible to enter into the study [9, 10, 11, 12, 13]. Children underwent a physician health examination and skin prick testing to a panel of 15 aeroallergens at yearly clinic visits. In addition, at each visit a parent was administrated a standardized questionnaire which inquired about respiratory symptoms including wheezing during the past year and other information.

Clinical Outcome

This study analyzed persistent wheezing, and SPT positivity to a panel of 15 aeroallergens at age three. Persistent wheezing was defined as two or more episodes of wheezing in the previous 12 months and reported wheezing at the most recent clinical exam or parental report of a physician-diagnosed asthma. SPT positivity to aeroallergen was defined as at least one positive SPT of the 15 tested aeroallergens determined at the year three clinic visit, including cat and dust mite allergens. A positive SPT was defined as the diameter of a wheal $\geq 3$ mm than the saline control.

Home Dust Collection and Measurement of Allergen Exposures

Home dust samples were collected at home visits at approximately age one. There were 674 dust samples collected from the infants’ primary activity room, mostly from the family or living room. Details about home dust collection have been described elsewhere [7, 26]. Cat (fel d1) and dust mite (der f1) allergen concentrations in home dust samples were analyzed by Enzyme-Linked ImmunoSorbent Assays (ELISA). A heterogeneous variance four parameter mixed effects model was used to estimate the dose-response
relationship of the standard curve in each ELISA. Then the estimated standard curve was used to calibrate home allergen levels. Geometric means and 95% confidence intervals of cat and dust mite allergens were 1.5 \( \mu \text{g/g} \) (0-153.8) and 0.3 \( \mu \text{g/g} \) (0-15.6), respectively. Among the 674 home dust samples, 8% of cat and 35% of dust mite allergens were below the lower limit of detection.

Other Exposure Definitions

All exposures other than the two aeroallergens were dichotomized, including race (African American/Non-African American), early (age one) dog ownership (yes/no), parental asthma (yes/no), visible mold (yes/no), at least one sibling (yes/no), cigarette smoke exposure during infancy (yes/no), and lower (LRC), upper (URC) respiratory conditions at three years old (yes/no). Children with one African American parent were classified as African-Americans. Otherwise, they were classified as non-African Americans. Among non-African American children, 98 percent of non-African Americans were Caucasians and only 2 percent were Asian, Pacific Islander or else. Mold exposure was defined as “visible mold” when having mold surface area \( \geq 0.2 \text{m}^2 \) during a home visual inspection before one year old of a child’s age. Cigarette smoke exposure was defined as having a household member who smoked at least one cigarette per day. LRC included croup, respiratory flu, cystic fibrosis, viral infection, bronchitis and pneumonia. URC included ear and sinus infection, tonsillitis, and strep throat.

3. Statistical Methods

The allergic clinical outcomes of this study were persistent wheezing and positive SPT to cat and dust mite at age three. Primary exposures investigated were home cat and dust
mite allergens during infancy. Each clinical outcome was analyzed by a multiple logistic regression, which included both dust mite allergen and cat allergen as independent variables/covariates. The reference group for each clinical outcome was different. The reference group in the wheezing analyses was those who had no persistent wheezing at age three. To avoid cross-reactivity among allergen skin sensitizations, the reference group in both SPT analyses was those who had negative SPTs to all aeroallergens. Thus, the total number of children included in each analysis varied because of the definition of the outcome, the selection of the corresponding reference group and missing information on either the outcome or any independent variable.

Independent variables were considered for inclusion in the regression models if they were previously identified as predictors of similar clinical outcomes or believed a priori to be possible confounders and/or effect modifiers of the relationships between clinical outcomes and exposure variables. Therefore, these were considered for model inclusion. Interactions between variables were investigated and retained in the final models if p≤0.15.

Cat and dust mite allergen were analyzed continuously and were log-transformed to approximate symmetry. Smoothed plots were produced from separate generalized additive models of each clinical outcome versus cat and dust mite allergens. These plots were used to model allergen levels in multiple logistic models. Allergen levels and corresponding outcomes with obvious inflation points/trend changes in the smoothed plots were modeled non-linearly. Several studies used thresholds based on the distributions of the allergens. More specifically, most of them divided the allergen levels into subsets according to distributions of the allergens, such as tertiles, quintiles or
deciles [2, 3, 4, 5, 7, 8, 14, 15]. This study used a linear spline technique [16]. Home allergen levels were divided into subsets based on smoothed plots. Within each subset, the relationship between the clinical outcome and home allergen level was modeled linearly and a trend difference in these linear relationships was allowed. For relationships without obvious inflation points, allergen levels were modeled with a constant linear trend across the entire range of allergens. The advantage of the linear spline approach was that not only differences between subsets were recognized but also the trend within each subset was quantified. The R statistical software package was used to analyze and produce plots of the data [17].

4. Results

The final logistic models of persistent wheezing included the following variables: cat allergen and dust mite allergen (model continuously), gender, race, visible mold, dog ownership, parental asthma, at least one sibling, cigarette smoke exposure, LRC, URC, and the interaction between SPT positivity to any aeroallergen and race. The same persistent wheezing model was applied for African American and Non-African American separately after removing race, the interaction of race and SPT positivity to any aeroallergen. Final SPT outcome models (cat and dust mite SPT positivity) included the same variables except LRC, URC and SPT positivity.
Figure 1. Smoothed plots of cat allergen and dust mite vs: the adjusted prevalence rates of persistent wheezing (a-b), and the adjusted prevalence of SPT + to cat (c-d) and dust mite (e-f).
Table 1. Descriptive statistics (frequencies) of independent variables and adjusted odds ratios (OR) with their 95% confidence intervals (CI) in persistent wheezing multiple logistic models.

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Model (1) Persistent Wheezing 71/525 (13.5%)</th>
<th>Model (2) Persistent Wheezing for African Americans 20/98 (20.4%)</th>
<th>Model (2) Persistent Wheezing for Non-African Americans 51/427 (11.9%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N (%)</td>
<td>OR (95% CI)</td>
<td>N (%)</td>
</tr>
<tr>
<td>Cat allergen*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 6.1 mg/g (low)</td>
<td>396 (75.4)</td>
<td>1.0 (0.8, 1.3) &amp;</td>
<td>89 (90.8)</td>
</tr>
<tr>
<td>≥6.1 mg/g (high)</td>
<td>129 (24.6)</td>
<td>0.7 (0.4, 1.3) &amp;</td>
<td>9 (9.2)</td>
</tr>
<tr>
<td>Dust mite allergen*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;0.2 mg/g (low)</td>
<td>264 (50.3)</td>
<td>0.8 (0.6, 1.0) &amp;</td>
<td>43 (43.9)</td>
</tr>
<tr>
<td>≥0.2 mg/g (high)</td>
<td>261 (49.7)</td>
<td>2.7 (1.1, 6.4) &amp;</td>
<td>55 (56.1)</td>
</tr>
<tr>
<td>Gender (male)</td>
<td>284 (54.1)</td>
<td>2.6 (1.5, 4.8) &amp;</td>
<td>56 (57.1)</td>
</tr>
<tr>
<td>Visible mold (yes)</td>
<td>296 (56.4)</td>
<td>1.8 (1.0, 3.3) &amp;</td>
<td>42 (42.9)</td>
</tr>
<tr>
<td>Dog Ownership (yes)</td>
<td>189 (36.0)</td>
<td>1.1 (0.6, 2.0)</td>
<td>9 (9.2)</td>
</tr>
<tr>
<td>Parental asthma (yes)</td>
<td>179 (34.1)</td>
<td>2.9 (1.7, 5.1)</td>
<td>44 (44.9)</td>
</tr>
<tr>
<td>Lower respiratory condition</td>
<td>148 (28.2)</td>
<td>2.4 (1.3, 4.3)</td>
<td>16 (16.3)</td>
</tr>
<tr>
<td>Upper respiratory condition</td>
<td>262 (49.9)</td>
<td>1.6 (0.9, 2.8)</td>
<td>38 (38.8)</td>
</tr>
<tr>
<td>Cigarette smoke exposure</td>
<td>124 (23.6)</td>
<td>2.3 (1.3, 4.3)</td>
<td>30 (30.6)</td>
</tr>
<tr>
<td>Sibling (yes)</td>
<td>417 (79.4)</td>
<td>0.7 (0.4, 1.4)</td>
<td>68 (69.4)</td>
</tr>
<tr>
<td>African American Race (yes)</td>
<td>98 (18.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aeroallergen SPT (+)</td>
<td>47 (48.0)</td>
<td>4.8 (1.4, 16.7)</td>
<td>47 (48.0)</td>
</tr>
<tr>
<td>Other Race (yes)</td>
<td>427 (81.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aeroallergen SPT (+)</td>
<td>174 (40.7)</td>
<td>1.4 (0.7, 2.5)</td>
<td></td>
</tr>
</tbody>
</table>

*Interpretation of ORs: for cat and dust mite allergen, ORs represent odds of wheezing when the allergen levels increase from the geometric mean to one geometric standard deviation above the geometric mean; for all other independent variables, ORs represent the odds of wheezing for the independent variable’s indicated category compared to the baseline (reference) category.
**Persistent Wheezing**

Figures 1(a) shows that early exposure to lower levels (below 6.1 µg/g) of cat allergen was weakly associated with the adjusted prevalence of persistent wheezing. When the exposure levels were high (above 6.1 µg/g), the curve showed that early exposure to cat allergen reduced the risk of developing persistent wheezing at age three. An opposite nonlinear relationship was observed in Figure 1 (b). Dust mite allergen was protective from having persistent wheezing at low levels (below 0.2 µg/g); however, it became a risk factor at high levels (above 0.2 µg/g). Based on these two figures, in multiple logistic regressions, cat allergen levels were divided at 6.1 µg/g, which was approximately the 75th percentile of this cohort, and dust mite allergen levels were divided at 0.2 µg/g, which was close to the median of this cohort.

In the multiple logistic regression (Table 1, Model 1), of the 525 children who had complete information of both clinical outcome and independent variables, 71 (13.5%) had persistent wheezing at age three. Only when home dust mite allergen level was above 6.1 µg/g, did the prevalence of persistent wheezing have a significant positive relationship with home dust mite allergen level (OR 2.7, 95% CI 1.1-6.4). Male gender (2.6, 1.5-4.8), visible mold (1.8, 1.0-3.3), parental asthma (2.9, 1.7-5.1), LRC (2.4, 1.3-4.3) and, cigarette smoke exposure (2.3, 1.3-4.3) were significant risk factors of persistent wheezing as well. A significant interaction effect revealed that African Americans and non-African Americans were different when they had positive SPT to aeroallergen. SPT positivity was only a risk factor for African Americans with an OR 4.8 (95% CI=1.4-16.7), which was almost four fold higher than for Non-African Americans.
Due to sensitivities of the ELISA assays, a portion of the home dust samples in the low exposure groups were not detectable and their allergen concentrations were replaced by the detection limits of the respective assays. Among 396 home dust samples with cat allergen <6.1 µg/g, 11.6% were below the LLOD. Among 264 home dust samples with dust mite allergen <0.2 µg/g, 67.8% were below the LLOD. Therefore, the relationship between allergen exposure and the clinical outcome could be distorted at lower levels of both allergens, especially dust mite allergen.

Among the 525 children included in the persistent wheezing analysis, African American children were more likely to have persistent wheezing at three years of age. Twenty out of 98 (20.4%) African Americans and 51 out of 427 (11.9%) Non-African Americans had persistent wheezing at age three (p=0.03). Figure 2 indicated that at three years old, African American children had a significantly higher prevalence of SPT positivity to aeroallergens, contributed by a significantly higher sensitization to cat, dog and mold.

To further investigate these racial differences, the same multiple logistic model was repeated for the 98 African American children and non-African American children separately (Table 1, Model 2). Parental asthma and LRC were risk factors for persistent wheezing in both races, but were much higher risks for African American children. Having a positive SPT to aeroallergen only increased the odds of having persistent wheezing for African American children. Early exposure to high levels of dust mite, male gender and cigarette smoke were only significant risks for non-African Americans.
Figure 2. Prevalence of positive SPTs among 98 African Americans children and 427 non-African American children.
Table 2. Demographic characteristics according to child’s race.

<table>
<thead>
<tr>
<th>Characteristic, N (%)</th>
<th>African Americans Total=98</th>
<th>Non-African Americans Total=427</th>
<th>P Value&lt;sup&gt;e&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cat allergen ≥6.1 μg/g, yes</td>
<td>9 (9.2)</td>
<td>120 (28.1)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Dust mite allergen ≥0.22 μg/g, yes</td>
<td>55 (56.1)</td>
<td>206 (48.2)</td>
<td>0.16</td>
</tr>
<tr>
<td>First year owned a dog, yes</td>
<td>9 (9.2)</td>
<td>180 (42.2)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>First year owned a cat, yes</td>
<td>5 (5.1)</td>
<td>113 (26.5)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Gender, male</td>
<td>56 (57.1)</td>
<td>228 (53.3)</td>
<td>0.50</td>
</tr>
<tr>
<td>Visible mold, yes</td>
<td>42 (42.9)</td>
<td>254 (59.5)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Parental asthma, yes</td>
<td>44 (44.9)</td>
<td>135 (31.6)</td>
<td>0.01</td>
</tr>
<tr>
<td>Sibling, yes</td>
<td>68 (69.4)</td>
<td>349 (81.7)</td>
<td>0.01</td>
</tr>
<tr>
<td>Cigarette smoke exposure, yes</td>
<td>30 (30.6)</td>
<td>94 (22.0)</td>
<td>0.07</td>
</tr>
<tr>
<td>Upper respiratory condition, yes</td>
<td>38 (38.8)</td>
<td>224 (52.5)</td>
<td>0.01</td>
</tr>
<tr>
<td>Lower respiratory condition, yes</td>
<td>16 (16.3)</td>
<td>132 (30.9)</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Family income ≤ 20,000, yes</td>
<td>45 (45.9)</td>
<td>25 (5.9)</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

<sup>e</sup> P values were from χ² statistics.
Table 2 showed the frequencies of wheezing related demographic characteristics (independent variables in Model 2 in Table 1) plus family income by racial status. The relationship between these characteristics and racial status were evaluated by univariate $\chi^2$ statistics. African American children were less likely to have a cat, high levels of home cat allergen, a dog at home, visible mold at home, a sibling, URC and LRC, but more likely to have a history of parental asthma and annual family income less than 20,000 dollars. Furthermore, among those who had persistent wheezing at age three, 10 out of 20 (50.0%) African American children used asthma medication and 38 of 51 (74.5%) Non-African American children used asthma medication. Controlling for the wheezing status, African American children were significantly less likely to use any asthma medication (Cochran-Mantel-Haenszel test, p=0.04).
Table 3. Descriptive statistics (frequencies) of independent variables and adjusted odds ratios (OR) with their 95% confidence intervals (CI) in SPT models.

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Model (2) SPT to Cat 44/351 (12.5%)</th>
<th>Model (2) SPT to Dust Mite 48/355 (13.5%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N (%)</td>
<td>OR (95% CI)</td>
</tr>
<tr>
<td>Cat allergen*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 6.1 mg/g (low)</td>
<td>262 (74.6)</td>
<td>1.4 (0.9, 2.0)</td>
</tr>
<tr>
<td>≥ 6.1 mg/g (high)</td>
<td>89 (25.4)</td>
<td>0.6 (0.3, 1.2)</td>
</tr>
<tr>
<td>Dust mite allergen*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 0.2 mg/g (low)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥ 0.2 mg/g (high)</td>
<td>351 (100)$</td>
<td><strong>0.6 (0.4, 0.9)</strong>$</td>
</tr>
<tr>
<td>Gender (male)</td>
<td>186 (53.0)</td>
<td>1.4 (0.7, 2.8)</td>
</tr>
<tr>
<td>Visible mold (yes)</td>
<td>189 (53.9)</td>
<td>1.1 (0.5, 2.2)</td>
</tr>
<tr>
<td>Dog Ownership (yes)</td>
<td>121 (34.5)</td>
<td><strong>0.3 (0.1, 0.8)</strong></td>
</tr>
<tr>
<td>Parental asthma (yes)</td>
<td>124 (35.3)</td>
<td><strong>2.6 (1.3, 5.2)</strong></td>
</tr>
<tr>
<td>Cigarette smoke exposure (yes)</td>
<td>89 (25.4)</td>
<td>0.8 (0.4, 1.8)</td>
</tr>
<tr>
<td>Sibling (yes)</td>
<td>278 (79.2)</td>
<td>0.6 (0.3, 1.4)</td>
</tr>
<tr>
<td>African American Race (yes)</td>
<td>67 (19.1)</td>
<td>2.0 (0.9, 4.6)</td>
</tr>
</tbody>
</table>

$ In Model (3) and (4), dust mite allergen was log transformed and modeled linearly. In Model (4), cat allergen was log transformed and modeled linearly.
SPT Sensitization to Cat and Dust Mite

In a previous study of this cohort, at age one of 680 infants with at least one atopic parent, the prevalence rate of SPT positivity to cat and dust mite were 2.5 and 3.1 percent respectively [30]. This study showed that at three years old of the remaining 662 children 7.8 percent had a positive SPT to cat and 8.6 percent had a positive SPT to dust mites. In the two allergen specific SPT outcome analyses, 12.5 percent had positive SPT to cat and 13.5 percent had positive SPT to dust mites.

Figure 1 (c) shows that the highest prevalence of persistent wheezing was around 6.1\mu g/g. At low levels of cat allergen, the prevalence of SPT positivity had an upward trend. The prevalence was increasing with the level of allergen exposure. At high levels of cat allergen, the trend was reversed. Figure 1 (d) shows that dust mite allergen levels had a negative association with the prevalence of SPT positivity to cat allergen across the entire range of measurable levels of dust mite allergen.

In a multiple logistic model (Table 3 Model 3), 351 children were included and 44 were skin sensitized to cat. The only significant exposure-response relationship was between the level of exposure to home dust mite and SPT positivity to cat (0.6, 0.4-0.9). In addition, owning a dog significantly reduced the risk of having positive SPT to cat allergen (0.3, 0.1-0.8). Having a parent with asthma increased the child’s likelihood of being SPT positive to cat (2.6, 1.3-5.2).

Figure 1 (e) and (f) shows the prevalence of SPT positivity to dust mite had flat relationships with both allergens. These weak associations were confirmed by the
multiple logistic model (Table 3 Model 4). Forty-eight of 355 children were included in
the analysis had a positive SPT to dust mites. No covariate significantly predicted SPT
positivity to dust mite allergen, although male gender (1.7, 0.9-3.3), visible mold (1.3,
0.7-2.5) and cigarette smoke exposure (1.3, 0.6-2.6) were found to be the as three highest
risk factors.

As fewer numbers of children were included in the allergen specific skin sensitization
analyses compared to the persistent wheezing analysis, subgroup analyses for African
Americans and Non-African Americans were not attempted.

5. Discussion

Home Allergens

Nonlinear exposure-response relationships between dust mite allergens and persistent
wheezing were shown in smoothed plots. The threshold 0.2 μg/g modified the
relationship between early exposure to dust mite allergen and persistent wheezing. When
dust mite allergen exposure was above this threshold, early exposure of dust mites
significantly increased the risk for developing persistent wheezing later in a child’s life,
especially to non-African American children. No association was found between dust
mite allergen and skin sensitization to dust mites. These two findings were consistent
with an Australian birth cohort study [14], even though their dust mite allergen levels
were a time weighted average of first five year visits. However, considering dust mites
are mostly related to the living environment and housekeeping habits of the family, which
should be very consistent during the first few years of child’s life, our results should be
comparable. Their study demonstrated that the time weighted-average dust mite allergen
levels of a child’s first five years of life had a significant positive association with asthma at age five years when the dust mite level was below 13.5 µg/g, while beyond 13.5 µg/g the prevalence of asthma started to drop. Their threshold (13.5 µg/g) was very close to the 95th percentile of dust mite allergen level in this CCAAPS cohort and their asthma definition was very similar to the persistent wheezing definition in this paper. This same birth cohort study also found dust mite allergen and mite sensitization had a significant negative association when dust mite levels were above 13.5 µg/g. This probably explained the reason that this CCAAPS cohort showed positive association between dust mite allergen and persistent wheezing but did not show any correlation between dust mite and mite sensitization, since very few children were exposed to dust mite allergen above 13.5 µg/g at one year old. Therefore, our findings indicated that avoidance of high levels of early dust mite allergen exposure could reduce the risk of developing persistent wheezing prevalence but probably not the risk of mite skin sensitization in a relatively low mite exposure environment like homes in the Greater Cincinnati area.

Very few studies investigated the cross reactivity between two different allergens. This paper found that dust mite allergen was significantly correlated to the prevalence of skin sensitization to cat allergen. This study found that increasing exposure of dust mite allergen was associated with a significantly lower risk of skin sensitization to cat at three years of age. This creates a paradox of early dust mite allergen control to reduce persistent wheezing and skin sensitization to cat. If considering it as a much serious immune response, instructions should be given to parents especially to Non-African Americans to avoid high levels of dust mite allergen. Increased vacuuming frequency and
use of a forced-air heating system have been associated with low mite concentrations by a previous study of this cohort [18].

In this study, early exposure to home cat allergen showed no significant association to any clinical outcome of children at age of three, even to sensitization to cat allergen. This finding is consistent with a population based cross sectional study of school children (age 12-14). They found a positive correlations between IgE antibodies to cat and exposure to cat allergen, but no exposure-response relationship was found between cat allergen and skin sensitization or asthma. [19]

**Race and Wheezing**

Several other studies reported that asthma and wheezing prevalence and the morbidity and mortality rates of African American children were disproportionately high [20, 21]. African Americans and Non-African Americans had different asthma symptoms [22], and different socioeconomic, environmental and genetic characteristics [23]. An early study of CCAAPS cohort demonstrated that among African Americans infants CT and TT genotypes of the IL-4 C-589T SNP significantly modified the effects of high ETS exposure on the development of wheezing without a cold at age one [24]. This study showed African American children had significantly higher prevalence of persistent wheezing and SPT positivity to aeroallergen. Compared to non-African American children African American children were more than two times more likely to have persistent wheezing when they had parental asthma, lower respiratory conditions or skin sanitization to aeroallergen, although these three factors were risks factors for both races. During the first year of life, living in a home with high level of dust mite allergen or
having one smoker in the household significantly increased a non-African American child’s risk of developing persistent wheezing at age three. Demographically, more African American children had a history of parental asthma and came from a low-income family. This study observed that among those three year old who had persistent wheezing only half of the African Americans used asthma medication, which was significantly less than non-African Americans. Similar findings were reported in three other studies indicating that minority asthma patients had less anti-inflammatory medication prescriptions [25, 26, 27]. All these findings indicate that it is important for physicians and other medical professionals to recognize these racial differences. Acknowledging how specific exposures affect different races, their specific symptoms and their specific socioeconomic and environmental characteristics can lead to more effective asthma management and patient care.

In summary, this CCAAPS cohort demonstrated that controlling early exposure to home dust mite allergen to a low level reduced children’s risks of developing persistent wheezing at age three. The levels of early exposure to cat allergen showed no association with any of three clinical outcomes at age three. Follow-up investigations are needed to determine the effects of early exposures to these two allergens on these children’s lives beyond the age of three, as this cohort study continues. African American and Non-African American American children showed differences in prevalence of persistent wheezing and skin sensitizations at the age of three, as well as in their different responses to exposures that probably contributed to their socioeconomic and environmental characteristics. Future studies should attempt to obtain broad characteristics of both races
and a larger sample of African American children, in order to understand the racial differences and provide better asthma patient care.
Bibliography


[14] Tovey ER, Almqvist C, Li Q, Crisafulli D, Marks DB. Nonlinear relationship of mite allergen exposure to mite sensitization and asthma in a birth cohort. American Academy of Allergy, Asthma & Immunology. 2008, 122, 114-118.


[23] Cloutier MM. Considerations in Culturally Directed Asthma Disease Management Programs. Disease Manage Health Outcomes 2008; 16 (2): 95-105


Chapter IV. Conclusion

This dissertation discusses a nonlinear mixed effect method to estimate the standard curve of immunoassay data in order to calibrate concentration levels in the analyzed samples. It also provides methods to calibrate sample concentrations and evaluate assays’ sensitivity and repeatability. These methods have been applied on dust mite, cat, dog and roach assays data in the CCAAPS study. Compared to the portions of samples which had been analyzed by a simple log-log linear method, this nonlinear method can estimate samples with concentrations close to the lower and upper asymptotes of the standard curves.

The dust mite and cat allergen concentrations from dust samples collected from infants home at approximately at age one The exposure-response relationships of several exposures and clinical allergic health outcomes at age three are investigated. Only first year dust mite allergen exposure showed effects on persistent wheezing and SPT positivity to cat allergen. First year exposure to cat allergen has no effects on persistent wheezing, SPT positivity to cat and SPT positivity to dust mite allergen. African American and non-African American are observed different in their demographic characteristics and prevalence of persistent wheezing and SPT positivity. These results should be acknowledged by health providers and parents to improve the quality of patient care.

The limitations of this dissertation are mentioned in the end of chapter I and chapter II. In terms of statistical methods of immunoassay analysis, more sophisticated goodness of fit tests should be explored to compare several linear and nonlinear models. Chapter III explores the risk/protective factors of the clinical allergic outcomes at age three in
CCAAPS cohort. Similar studies should be repeated when the same outcomes are available at later ages. The results of chapter III are limited by the population of CCAAPS cohort. A meta-analysis can be conducted to pool the information from several studies/populations in order to get a wider range of allergen concentrations.
## Supplements and Appendices

### Supplemental Table and Figure

Table S1. A Dust Mite Assay Standard Data with Ten Standard Concentration Levels

<table>
<thead>
<tr>
<th>Plate # Standard Concentration (µg/ml)</th>
<th>Plate 1</th>
<th>Plate 2</th>
<th>Plate 3</th>
<th>Plate 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>250.00</td>
<td>OD1 2.38 OD2 2.60</td>
<td>OD1 2.42 OD2 2.33</td>
<td>OD1 2.40 OD2 2.21</td>
<td>OD1 2.25 OD2 2.27</td>
</tr>
<tr>
<td>125.00</td>
<td>OD1 2.36 OD2 2.43</td>
<td>OD1 2.25 OD2 2.35</td>
<td>OD1 2.49 OD2 2.33</td>
<td>OD1 2.36 OD2 2.42</td>
</tr>
<tr>
<td>62.50</td>
<td>OD1 2.29 OD2 2.25</td>
<td>OD1 2.04 OD2 2.25</td>
<td>OD1 2.39 OD2 2.33</td>
<td>OD1 2.28 OD2 2.18</td>
</tr>
<tr>
<td>31.25</td>
<td>OD1 1.91 OD2 1.93</td>
<td>OD1 1.87 OD2 1.80</td>
<td>OD1 1.81 OD2 1.84</td>
<td>OD1 1.66 OD2 1.69</td>
</tr>
<tr>
<td>15.63</td>
<td>OD1 1.30 OD2 1.34</td>
<td>OD1 1.34 OD2 1.38</td>
<td>OD1 1.34 OD2 1.36</td>
<td>OD1 1.32 OD2 1.33</td>
</tr>
<tr>
<td>7.81</td>
<td>OD1 0.79 OD2 0.84</td>
<td>OD1 0.71 OD2 1.01</td>
<td>OD1 0.77 OD2 0.83</td>
<td>OD1 0.74 OD2 0.73</td>
</tr>
<tr>
<td>3.91</td>
<td>OD1 0.38 OD2 0.42</td>
<td>OD1 0.36 OD2 0.48</td>
<td>OD1 0.40 OD2 0.42</td>
<td>OD1 0.40 OD2 0.46</td>
</tr>
<tr>
<td>1.95</td>
<td>OD1 0.20 OD2 0.22</td>
<td>OD1 0.18 OD2 0.24</td>
<td>OD1 0.21 OD2 0.21</td>
<td>OD1 0.19 OD2 0.22</td>
</tr>
<tr>
<td>0.98</td>
<td>OD1 0.10 OD2 0.11</td>
<td>OD1 0.09 OD2 0.12</td>
<td>OD1 0.12 OD2 0.11</td>
<td>OD1 0.11 OD2 0.10</td>
</tr>
<tr>
<td>0.49</td>
<td>OD1 0.05 OD2 0.05</td>
<td>OD1 0.06 OD2 0.06</td>
<td>OD1 0.07 OD2 0.06</td>
<td>OD1 0.04 OD2 0.05</td>
</tr>
</tbody>
</table>
Figure S1. Supplemental Graphical Display of Standard Curve Data from Four Plates in the Dust Mite Assay shown in Table S1.
Appendix I. Iteratively Re-weighted Pseudo-likelihood Estimation

The estimates of $\beta$, $\sigma$ and $\theta$ are obtained simultaneously using iteratively re-weighted pseudo-likelihood estimation. The estimation process is described as followed.

(a). Estimate $\beta$ through Ordinary Least Square (LS). i.e. assume $\theta=0$ and $\mu_j=1$.

(b). Use the residuals to estimate variance parameters, $\sigma$ and $\theta$ in (2).

(c). Use the variance parameter estimates from step (b) to update $\beta$s through Weighted Least Square (WLS) with weights $w_j = \frac{1}{\hat{\sigma}^2 g^2(\hat{\mu}_j, \hat{\theta})}$.

(d). Repeat step (b) to (c) until convergence criteria are met.

PROC NLIN in Statistical software, SAS, can accomplish the steps listed above.
Appendix II. Testing Homogeneity of Intra-plate Variance

The assumption of homogeneity of intra-plate variance can be tested through two Chi-squared statistics (Zeng, 1997). Both statistics are based on the assumption that given the true parameters, intra-run parameter estimates are asymptotically normal, \( \hat{\eta}_i \mid \eta_i \sim N(\eta_i, C_i) \), where \( \eta \) is \((\log \sigma, 0)^t\).

The two Chi-squared statistics are

\[
Q_1 = \sum_{i=1}^{m} (\hat{\eta}_i - \bar{\eta}) C_i (\hat{\eta}_i - \bar{\eta}) \sim \chi^2(q(m-1)) ,
\]

\[
Q_2 = \sum_{i=1}^{m} (\hat{\eta}_i - \hat{\eta}_{pooled}) C_i (\hat{\eta}_i - \hat{\eta}_{pooled}) \sim \chi^2(m-1) .
\]

where \( m \) is the total number of plates within the same assay and \( q \) is the number of variance parameters. \( q \) is equal to 2 here. In \( Q_1 \), \( \bar{\eta} \) is the weighted average of \( m \) \( \hat{\eta} \)'s and

\[
\bar{\eta} = (\sum_{i=1}^{m} C_i^{-1})^{-1}(\sum_{i=1}^{m} C_i \hat{\eta}_i) .
\]

In \( Q_2 \), \( \hat{\eta}_{pooled} \) is the pooled variance estimates when pooling all plates within the same assay. The covariance matrix \( C_i \) is obtained from 500 bootstrap samples. For each plate, 500 bootstrap samples are created. 500 pairs of bootstrap variance estimates \( \hat{\eta} \) are obtained and \( C_i \) is calculated based on these 500 pair estimates. When the resulting p values from \( Q_1 \) and \( Q_2 \) are above 0.05, the homogeneous variance hypothesis is not rejected.
Appendix III. Lower Limit of Detection (LLOD) and Upper Limit of Detection (ULOD) Calculation

The one sided t test below is used to test the equality of the expected OD at the LLOD and the expected OD at zero concentration.

\[ P((\bar{y}(\hat{x}_{LLOD},\hat{\beta}) - \bar{y}(0,\hat{\beta}) \geq 0) \geq 1 - \alpha \text{ and } (\bar{y}(\hat{x}_{LLOD},\hat{\beta}) - \bar{y}(0,\hat{\beta})) \sim t_{\alpha, df} . \]

Here \( \bar{y}(x) \) is the expected OD from repeated measurements of ODs at concentration \( x \); \( \alpha \) is the probability that the expected OD at the LLOD is tested to be significantly higher than the expected OD at zero concentration when indeed two expected ODs are not significantly different, i.e. the type I error rate or the false positive rate. Most often \( \alpha \) is chosen to be 0.05.

LLOD values are plate-specific and are calculated as follows: (Davidian, 1995).

\[ \{ f(\hat{x}_{LLOD},\hat{\beta}) - f(0,\hat{\beta}) \} = t_{\alpha, df} \sqrt{[\sigma^2 g^2 ((f(\hat{x}_{LLOD},\hat{\beta}),\hat{\theta})I + f_{\hat{\beta}}(0,\hat{\beta})\hat{\Sigma}(\hat{\beta})f_{\hat{\beta}}(0,\hat{\beta})]}. \tag{7} \]

The left hand side of the equation is the difference between the OD at the LLOD and the OD at zero concentration on the standard curve. The right hand side of the equation is the product of a t statistic and the standard deviation of the left hand side of the equation. The parameter estimates, \( \hat{\beta}, \hat{\Sigma}(\hat{\beta}) \) and \( \hat{\theta} \), in the formula are obtained from the NLME model; \( \hat{\Sigma}(\hat{\beta}) \) is the covariance of \( \hat{\beta} \). The number of degrees of freedom of the t distribution is \( (N - 4m) \), where \( N \) is the total number of observations in the assay, and \( m \) is the number of plates in the assay.

The ULOD is the concentration close to the saturation of the assay. The ULOD is obtained from \( \hat{\beta}_i \) in the FPL model.

\[ \hat{\beta}_i - f(\hat{x}_{ULOD},\hat{\beta}_i) = t_{\alpha, df} \sqrt{[\sigma^2 g^2 ((f(\hat{x}_{ULOD},\hat{\beta}_i),\hat{\theta}) + \sigma^2 g^2 (\hat{\beta}_i,\hat{\theta})]}. \tag{8} \]
Appendix IV. Assessing Experimental Conditions which Affect Dog Assay Variability

Table A.1. Experimental Conditions Which are Varied in Monoclonal Dog Allergen Assays

<table>
<thead>
<tr>
<th>Run #</th>
<th>Assay Date</th>
<th>number of plates in the run</th>
<th>ABTS (9/29/05)</th>
<th>ABTS (3/8/06)</th>
</tr>
</thead>
<tbody>
<tr>
<td>48</td>
<td>12/28/2005</td>
<td>2</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>50</td>
<td>1/20/2006</td>
<td>2</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>51</td>
<td>3/8/2006</td>
<td>4</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>54</td>
<td>2/23/2006</td>
<td>4</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>55</td>
<td>3/8/2006</td>
<td>4</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>56</td>
<td>3/23/2006</td>
<td>3</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>57</td>
<td>3/28/2006</td>
<td>4</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>58</td>
<td>4/7/2006</td>
<td>1</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>64</td>
<td>9/20/2006</td>
<td>3</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>65</td>
<td>9/26/2006</td>
<td>3</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>66</td>
<td>10/5/2006</td>
<td>1</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>67</td>
<td>11/14/2006</td>
<td>1</td>
<td>no</td>
<td>yes</td>
</tr>
</tbody>
</table>

Twelve runs of monoclonal dog allergen assays have records of their experimental conditions. Among them, only the date of the ABTS and the number of plates in the run are varied across runs. ABTS is the chromogenic substrate for peroxidase, which yields a yellow color measured by the plate reader. In Table A.1, ABTS (9/29/05) and ABTS (3/8/06) columns indicate the date of the ABTS. A mixed effects model was used to test the effects of the experimental conditions on the variability of the parameter estimates $\beta_1$ through $\beta_4$. The two experimental conditions, the date of the ABTS and the number of plates in the run, were modeled as fixed effects, and plate nested within run was treated as a random effect. For each parameter estimate, there were 32 observations. Neither the date of the ABTS nor the number of plates in the run had a significant effect on any of the parameter estimates $\beta_1$ through $\beta_4$. This means that variation in the experimental conditions did not have an effect on the shape of the mean standard curve. Since sample concentrations are calibrated based mainly on the mean standard curve, variation in the experimental conditions will not have an influence on the magnitudes of the calibrated sample concentrations.
Effects of the experimental conditions on the variance parameter estimates were also tested under the same mixed effects model. The dependent variables were the logarithm of $\theta$ and the logarithm of $\sigma$. No significant effects were found. Therefore, the experimental condition ABTS has no effect on the variance of the assays. Furthermore, it has no impact on the assay system in terms of the FPL parameter estimates from NLMIXED.

Since the sensitivity of the assay is very important, effects of the date of ABTS and the number of plates in the run on the estimated LLODs are also assessed using the same model. 32 estimated plate-specific LLODs were estimated from the 12 runs. Both the date of the ABTS and the number of plates in the run have significant effects on the estimated LLOD. The mean LLOD for runs using ABTS ordered at 9/29/05 is 1.98 ng/ml. The mean LOD for runs using ABTS ordered at 3/8/06 is 2.65 ng/ml. The runs with only one or two plates on average have a higher LLOD than runs with three or four plates. The average LLOD for runs with one plate is 2.96 ng/ml, for runs with two plates is 2.80 $\mu$g/ml, for runs with three plates is 1.57 $\mu$g/ml, and for runs with 4 plates is 1.92 $\mu$g/ml.

Figure A1 Mean LOD of ABTS Ordered at 9/29/05 and at 3/8/06, with Their Standard Error Bars
Mean LOD by ABTS Order Date

ABTS Order Date

3/8/06  9/29/05
Based on this analysis of the effects of experimental conditions on the estimated LLODs obtained from 12 monoclonal dog assays, there are two conclusions: (1) ABTS ordered at 3/8/06 might have an adverse impact on the LLOD. Due to an absence of other information about the experimental conditions, it is not clear whether the effect is due to the ABTS itself or some other condition which is confounded with ABTS. (2) By increasing the number of plates in the run, the assay became more sensitive to low level of concentrations in samples. Hence, controlling experimental conditions is crucial in assay analysis. Even though slight variations in
experimental conditions may not affect the magnitude of the estimated sample concentrations through the estimated regression parameters in the FPL model, variations in experimental conditions may have a significant effect on the sensitivity of the assays (i.e., the estimated LLODs). This will in turn cause challenges in estimating low level sample concentrations.
Appendix V. Comparison of Different Logistic Models

The usual logistic regression model is

\[ y = \frac{1}{1 + \exp(-\beta(x - \mu))}, \quad (A1) \]

where \( \mu \) denotes the median effective dose EC50 and the response \( y \) is bounded in \([0,1]\).

The FPL model is a more general form of logistic regression model, which is widely used in the bioassay literature. The four parameters in the FPL model have specific meanings and can assist in understanding the assay results. The original FPL model proposed by Rodbard and Hutt (1974) was

\[ y = d + \frac{a - d}{1 + (x / c)^{-b}} \] \quad (A2)

In this equation, \( y \) is the response, \( x \) is the dose, \( a \) and \( d \) are the responses at the lower and upper asymptotes, respectively, \( c \) is the expected dose which is half way between \( a \) and \( d \) and \( b \) is the slope of the linear portion of the curve. The slope is usually negative in radioligand assays and positive in allergen immunoassays. A2 can be written as

\[ \log it\left(\frac{Y - a}{d - a}\right) = b(\log x - \log c) \quad (A3a). \]

The logarithmic transformation of the dose \( x \) gives the more familiar logistic model. Clearly, in A3a the four parameters, \( a, b, c, d \), can be replaced by \( \beta_1, \beta_2, \beta_3 \) and \( \beta_4 \) in (f2.1),

\[ \log it\left(\frac{Y - \beta_1}{\beta_2 - \beta_1}\right) = -\beta_1(\log x - \beta_3). \quad (A3b) \]

One parameter, two parameter, or three parameter logistic functions are restricted models of the FPL model, when one or more parameter(s) are restricted to some specified value(s).
This is shown in Figure A1. Figure A1a is a dose response curve with four parameters: $\beta_1 = 2.5$, $\beta_2 = 0.2$, $\beta_3 = 2.5$ and $\beta_4 = 1.2$. The parameter values are chosen based on the example dust mite assay in Table 2.1. In Figure A1b, when $\beta_1 = 1$, $\beta_2 = 0$, and $\beta_3 = 0$ and $\beta_4 = 1.2$ or both $\beta_2$ and $\beta_3$ are restricted, function $f2.1$ becomes a two parameter logistic model, with the response bounded by $[0, 1]$ and no observable flat portion at the lower asymptote. In Figure A1c, when $\beta_1 = 1$, $\beta_2 = 0$, and $\beta_3 = 2.5$ and $\beta_4 = 1.2$ or only $\beta_2$ is restricted, function $f2.1$ becomes a three parameter logistic model. Due to different restrictions on the parameters in Figure A3a, A3b and A3c, the three curves in these figures have different shapes. Figure A1c has very similar shape as Figure A1a. The only difference is the intercept in Figure A1c is restricted to be 0. The response $y$ is bounded by $[0.2, 2.5]$, $[0,1]$ and $[0, 2.5]$, respectively. Figure A1a and A3c have the flat portion at both lower and upper asymptotes, which is often true for many assay systems especially those with wide dose ranges, such as the allergen assays in this paper. However, some assay systems may not have the symmetrical and S shaped dose response curve. With some further information of the assay systems, one or more parameter(s) in the FPL model can be set to some theoretical value(s). For example, in Figure A1b this particular assay system is assumed that it has no flat portion at the lower asymptote. As for any parametric regression model, the restrictions on the parameters in the FPL model can be tested by goodness of fit tests.

Figure A3. A1a is the plot from a FPL model, A3b is the plot from a two parameter logistic model, and A3c is the plot from a three parameter logistic model.
Appendix VI. Statistical Software

The statistical programs used in this paper are based on several SAS macro programs and one R program including standard curve estimation, testing homogeneous variance assumption across plates, precision profile and limits of detection calculation, and calibration of unknown samples. The standard curve estimation is based on FPL nonlinear mixed modeling with weighting. Estimates of the mean functions and the pooled variance parameter estimates of the assay can be obtained from the FPL model and are used in unknown sample calibration. Since a SAS macro is employed, users only need to enter a simple list of information, such as the name of the assay, number of plates in the assay run, dilution factor in each plate, etc. Therefore, users with limited SAS programming background can easily master the programs and analyze their assay data. However, prior the analysis, data have to be converted to SAS dataset format.

The statistical objective of any software for bioassay data should be to reduce bias of estimated samples and enhance assay precision. As mentioned above, statistical analysis of assay data includes two essential phases, standard curve estimation and sample calibration. At the same time, it is desirable to have the precision profile and LODs available. Precision profiles and LODs help users to understand the assay performance and can be used in further optimizing the experiment. In addition, to allow a successful analysis of the data, the software should accompany a data management function to help users in entering and modifying the data. At the same time a comprehensible manual with explanation of statistical methodologies should be accessible to users.

There are several commercial and free statistical software programs that claim that they can analyze assay data. Some of them are specifically designed for assay data, which can
calculate the sample concentrations based on the fitted curve and some key statistics to
evaluate the assay performance. Many statistical software programs only can do curve
fitting. These programs can fit a wide variety of linear and nonlinear curves. However,
additional software or at least additional programming is still necessary, in order to get
the calibrated concentrations for unknown samples. A brief survey is made to assess the
features, usability and accessibility of some of these software programs. Users should pay
close attention to the statistical methodologies in order to get appropriate and desirable
interpretation of the results.

R has a drc package (Ritz and Streibig, 2007). The advantage of drc from R is it is can be
downloaded from the internet without charge and it is maintained by a developers’ team.
This software can fit a variety of nonlinear models including three parameter, four
parameter, five parameter and modified four parameter logistic models and also a
Weibull model. It can fit an individual dose response curve and fit several dose response
curves simultaneously. However, the error of the response $\varepsilon$ is assumed to have an iid
$N(0, \sigma^2)$ distribution. Allergen assay data usually has a heterogeneous response error
distribution. Therefore, the precision and the sensitivity of an assay, which depend on the
estimates of the response error, may not be reliable, even though the mean function of the
curve is acceptable. Besides the rigid error distribution assumption, drc requires
familiarity with the R command syntax. Furthermore, drc doesn’t have a calibration
procedure. It is only a curve fitting program. Calibration needs be done by writing
additional R commands after extracting the fitted parameters of the dose response curve.

StatLIA from Brendan Technologies is a point and click statistical software package. It
is more appropriate for users with a limited programming background, as it does not
require the user to write any commands. It has 23 curve fitting options, including a 5 parameter logistic model with weighting, a 4 parameter logistic model with weighting, a logit-log model with weighting, a cubic spline model, and a linear model. StatLIA automatically provides the reportable range, based on the precision profile and the EC50 of each assay.

The 5 parameter logistic model is a modified 4 parameter logistic model. It adds an additional parameter \( \beta_5 \) \((\beta_5 > 0)\) representing the degree of the asymmetry, in order to accommodate the asymmetry of the dose response curve (Gottschalk and Dunn, 2005).

\[
y_j = f(x_j, \beta) = \beta_1 + \frac{(\beta_2 - \beta_1)}{1 + \exp(\beta_1 (\log x_j - \beta_3))} e_j.
\]

(A4)

In StatLIA the weighting factor is the inverse of the variance at each response level. The variance function is described by Gottschalk and Dunn (2005).

\[
\text{variance} = A(\text{response})^\theta
\]

(A5)

\( A() \) is a function of the averaged response and the averaged noise. \( \theta \) is the power of the \( A() \). In StatLIA, \( \theta \) is fixed at some predetermined level, which requires knowledge about the error distribution of the assay or historical data from the same assay technology.

Another drawback of StatLIA is that it doesn’t allow one to fit several individual assays simultaneously. Hence, the reliability of the variance estimates is doubtful.

**TableCurve2D** from SYSTAT is another point and click curve fitting software program. It includes numerous equations from linear to nonlinear (such as FPL). It can fit all possible equations on the same data and pick the best model based on goodness of fit statistics. It doesn’t allow heterogeneous variances. Weighted least square estimation is
not an option. Additionally, multiple assays can not be estimated simultaneously. Since it
is not a software program designed for assay data, no calibration can be done by the
software. Parameter estimates have to be extracted. Some other software has to be used
to calculate the precision profiles, limits of detection, and unknown sample
concentrations.

**Seelva** from Arlenda Laboratory Solution is a web based program designed for assay
data. It uses SAS software as its underlying program. Users do not need SAS to use
Seelva. However, internet access is required. The software can be accessed by entering a
valid user name and password. Like other point and click software programs, data have to
be entered in the required format. Many problems can stop the program, such as missing
values. The biggest drawback of this software in terms of statistical methodology is it
does not have a nonlinear curve fitting function. The best it can do is a weighted
quadratic regression model.
Appendix VII. R \textit{nlme} Random Effect Covariance Modeling

R \textit{nlme} package is employed to execute the NLME model in this paper. In R \textit{nlme} package, several possible random effect positive definite covariance structures can be chosen. They include

(1) The usual symmetric covariance structure with no additional restrictions, \textit{pdSymm} and \textit{pdLogChol}. For a 4x4 matrix, this covariance matrix contains 10 unique parameters.

(2) A diagonal matrix, \textit{pdDiag}. For a 4x4 matrix, this matrix only contains 4 unique parameters.

(3) Some other structures such as the identity, compound symmetry and blocked matrix also are available in \textit{nlme}. However, these covariance structures are not suitable for bioassay random effect models.

The choice of matrix (1) or (2) usually should be based on each specific dataset. (1) is always preferred with no additional knowledge of the covariance structure. When the covariance matrix is recognized as over-parameterized, a simpler matrix structure is recommended, for example (2) a diagonal covariance matrix structure may be chosen.
Appendix VIII. Results Comparison: Results from NLME Model vs. Results from the Log-log Model

In Table A.1, the results include 1526 dust mite allergen samples and 1524 cat allergen samples. These samples were analyzed by the method described in the paper (the NLME model). Among them, 28% of the samples’ dust mite allergen levels were below the LLOD, and 8% of the samples’ cat allergen levels were below the LLOD. No sample has a dust mite or cat allergen level above the ULOD. In Table A.2, the results include 1521 dust mite allergen samples and 1516 cat allergen samples. These samples were analyzed by using a Log-log model. Among them, about 61% of the samples’ dust mite allergen levels were below the LLOD or above the ULOD, and 28% of the samples’ cat allergen levels were below the LLOD or above the ULOD.

Among all the samples, 579 were within dust mite allergen limits of detection by using the NLME and the log-log model. The estimated correlation of logarithmic-transformed dust mite allergen levels from both methods is 0.88. 1092 sample concentrations were within the cat allergen limits of detection by using the NLME model and the log-log model. The estimated correlation of logarithmic-transformed cat allergen levels from both methods is 0.94.

Comparing Table A.1 and A.2, the LLODs from the NLME model are consistently lower than those from the log-log model. By using the FPL model, no samples’ dust mite or cat allergen levels are above the ULOD. Overall, using the NLME model can reduce the number of samples that have undetectable and/or inestimable concentrations. Therefore, using NLME to obtain allergen concentrations in environmental samples will increase the power of any statistical hypothesis test in health outcome-exposure studies.
Table A.1. Geometric Means and 95% Confidence Intervals (CIs) of Allergen Assays (Analyzed by NLME model)

<table>
<thead>
<tr>
<th></th>
<th>Geometric Mean Concentration (95 % CI*)</th>
<th>Geometric Mean LLOD (95%CI*)</th>
<th>Percent Below LLOD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dust Mite</td>
<td>0.40(0.03, 30.48) µg/g</td>
<td>0.05(0.02, 0.10) µg/g</td>
<td>28%</td>
</tr>
<tr>
<td>Cat</td>
<td>1.50(0.03, 272.76) µg/g</td>
<td>0.04(0.02, 0.13) µg/g</td>
<td>8%</td>
</tr>
</tbody>
</table>

*The 95% CI are calculated by percentile method.

Table A.2. Geometric Means and 95% Confidence Intervals (CIs) of Allergen Assays

(Analyzed by Log-log model)

<table>
<thead>
<tr>
<th></th>
<th>Geometric Mean Concentration (95 % CI*)</th>
<th>Geometric Mean LLOD (95%CI*)</th>
<th>Geometric Mean ULOD (95%CI*)</th>
<th>Percent Below LLOD and Above ULOD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dust Mite</td>
<td>0.47(0.10, 27.76) µg/g</td>
<td>0.19(0.10, 0.80) µg/g</td>
<td>58.33(50.00, 100) µg/g</td>
<td>60% and 1%</td>
</tr>
<tr>
<td>Cat</td>
<td>1.63(0.03, 250.00) µg/g</td>
<td>0.08(0.02, 0.50) µg/g</td>
<td>213.10(125.00, 500.00) µg/g</td>
<td>15% and 13%</td>
</tr>
</tbody>
</table>

*The 95% CI are calculated by percentile method.
Appendix IX. Comparison of Parameter Estimates from the NLME Model

From Table A.3, comparing the parameter estimates from the NLME models for dust mite allergen and cat allergen, the mean estimates and 95% CIs of the fixed effects $\beta_2$, $\beta_3$, and $\beta_4$ are very similar, but cat allergen has a significantly higher mean estimate with a significantly wider 95% CI of $\beta_1$. Therefore, when comparing the dust mite allergen standard curves, the cat allergen standard curves would be more likely to detect high concentration levels in the samples. Estimates of $\theta$ and $\sigma$ from cat allergen standard curves are consistently higher than estimates from dust mite allergen standard curves. This means that the cat allergen assays’ within plate variation is greater than the dust mite allergen assays’ within plate variation. From the mean standard curves of $\beta_1$, $\beta_2$, $\beta_3$, and $\beta_4$, cat allergen standard curves have smaller random effects. This means that the cat allergen assays’ between plate variation, on average, is smaller than the dust mite allergen assays’ between plate variation.
Table A.3. Mean and 95% Confidence Intervals (CIs) of Parameter Estimates From NLME

<table>
<thead>
<tr>
<th></th>
<th>Mean $\beta_1$ (95% CI*)</th>
<th>Mean $\beta_2$ (95% CI*)</th>
<th>Mean $\beta_3$ (95% CI*)</th>
<th>Mean $\beta_4$ (95% CI*)</th>
<th>Mean $\theta$ (95% CI*)</th>
<th>Mean $\sigma$ (95% CI*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dust mite</td>
<td>5.71</td>
<td>0.02</td>
<td>4.29</td>
<td>1.41</td>
<td>0.37</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>(1.81,49.63)</td>
<td>(0.008)</td>
<td>(2.58, 7.61)</td>
<td>(1.08, 1.95)</td>
<td>(-4.31, 1.12)</td>
<td>(0, 0.18)</td>
</tr>
<tr>
<td>Cat</td>
<td>25.16</td>
<td>0.02</td>
<td>3.13</td>
<td>1.78</td>
<td>0.75</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>(1.58, 800.83)</td>
<td>(-0.02, 0.11)</td>
<td>(1.03, 9.71)</td>
<td>(1.12, 2.50)</td>
<td>(0.18, 1.33)</td>
<td>(0.04, 0.43)</td>
</tr>
<tr>
<td>Random</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dust mite</td>
<td>0.90</td>
<td>0.01</td>
<td>0.24</td>
<td>0.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0, 14.70)</td>
<td>(0, 0.04)</td>
<td>(0, 0.88)</td>
<td>(0, 0.31)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cat</td>
<td>0.10</td>
<td>0.08</td>
<td>0.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0, 0.67)</td>
<td>(0, 0.02)</td>
<td>(0, 0.38)</td>
<td></td>
<td></td>
<td>(0, 0.14)</td>
</tr>
</tbody>
</table>

Note:

*The 95% CI are calculated by percentile method.

Std of $b_1$-$b_4$ (standard deviations of $b_1$-$b_4$): The squared stds of $b_1$-$b_4$ are the variances of $b_1$-$b_4$. They are the diagonal parameters in the random effect matrix. The fixed effects parameters $\beta_1$ – $\beta_4$ and the parameters from the random effect matrix help to understand the characteristics of the assay system.
Appendix X. Comparison of Several Standard Curve Fitting Approaches

Five standard curve fitting methods are compared based on their Sum of Squares of Errors and $R^2$. The data are in Table S1. The five methods are (1) least square linear regression, (2) least square cubic polynomial regression, (3) three parameter logistic regression with constant variance, (4) FPL model with constant variance, (5) FPL with heterogeneous variance. In all these models, the independent variable, $x$, is the logarithm of the concentration and the response variable, $y$, is the optical density. The $R^2$ from the log-log linear regression for all four plates is lower than any of the five models and is not shown here. For all five candidate models, the sum of squared errors is defined as the sum of squares of the deviation of the observed $y$ from the model predicted $y$. $SSE = \sum_{j=1}^{n} (y_j - \hat{y}_j)^2$. $\hat{y}_i$ is the predicted $y$ from the regression model at the corresponding concentration. $R^2$ is the fraction of the total sum of squares explained by the model. $R^2 = 1 - \frac{SSE}{SST}$, where $SST = \sum_{j=1}^{n} (y_j - \bar{y})^2$. $\bar{y}$ is the grand mean of all $y_j$.

From Table A1, the logistic models (model 3 and 4) have a better fit with a higher $R^2$ and lower SSE compared to the linear model and cubic model (model 1 and 2). Among the logistic models, the heterogeneous FPL model (model 5) has the highest $R^2$ for two plates from the example dust mite assay. For the other two plates, the three-parameter logistic model (model 3) has the best $R^2$ values. However, from the statistical methods described, it is important to allow the lower asymptote to vary and to estimate a lower asymptote ($\beta_2$) that is used to calculate the LLOD. For each plate, an F test is performed to test whether the three-parameter logistic model (model 3) is superior to the FPL (model 4). The P values from the F statistics are 0.181, 0.355, 0.156 and
0.384. Therefore, the three-parameter logistic model is not statistically different from the four-parameter logistic model.

R² is a popular goodness of fit statistics for linear regression models. It has limitations when it is applied on nonlinear regression models. R² of nonlinear regression models can lie outside [0, 1] and may decrease as regressors are added (Cameron, 1997). For all plates in dust mite assays and in cat assays, omega squared (ω²) and epsilon squared (ε²) are calculated to assess the goodness of fit of heterogeneous four parameter logistic models (Winer et al., 1991). ω² and ε² are “shrunken” goodness statistics which take account of the number of parameters in the model.

\[
\omega^2 = \frac{SS_{model} - (k - 1)MS_{error}}{SS_{total}}
\]

\[
\varepsilon^2 = \frac{SS_{treat} - (k - 1)MS_{error}}{SS_{total} + MS_{error}}
\]

In the formula, all Sum of Squares (SS_{total}, SS_{treat} and SS_{error}) are weighted by the variances of predicted responses. In Table A5, ω² and ε² of 129 dust mite plates and 128 cat plates were calculated. The median ω² and ε² are all greater than 0.99. Therefore, the four parameter logistic model with heterogeneous variances is a sufficient model to describe both dust mite allergen data and cat allergen data.
Table A4. Comparison of Standard Curve Models by Using Sum Squared Error and $R^2$

<table>
<thead>
<tr>
<th>Plate 1</th>
<th>Model</th>
<th>Linear</th>
<th>Cubic</th>
<th>Three Parameter Logistic</th>
<th>Four Parameter Logistic with Constant Variance</th>
<th>Four Parameter Logistic with Heterogeneous Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimated</td>
<td>y = 0.011+ 0.049<em>x+ 0.250</em>x^2 - 0.092*x^3</td>
<td>y = 2.559- 2.559/(1+exp(1.302*(x-2.645)))</td>
<td>y = 2.539+(0.0432 - 2.539)/(1+exp(1.269*(x-2.699)))</td>
<td>0.794</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td>y = 0.076+ 0.467*x</td>
<td>y = 2.539+(0.0432 - 2.539)/(1+exp(1.269*(x-2.699)))</td>
<td>0.794</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sum of Errors of Errors</td>
<td>0.901</td>
<td>0.116</td>
<td>0.044</td>
<td>0.039</td>
<td>0.027</td>
<td></td>
</tr>
<tr>
<td>R^2</td>
<td>0.950</td>
<td>0.994</td>
<td>0.999</td>
<td>0.998</td>
<td>0.998 (&gt;0.999*)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Plate 2</th>
<th>Model</th>
<th>Linear</th>
<th>Cubic</th>
<th>Three Parameter Logistic</th>
<th>Four Parameter Logistic with Constant Variance</th>
<th>Four Parameter Logistic with Heterogeneous Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimated</td>
<td>y = 0.033+ 0.085<em>x+ 0.224</em>x^2 - 0.030*x^3</td>
<td>y = 2.435-2.435/(1+exp(1.255*(x-2.556)))</td>
<td>y = 2.427+(0.017 - 2.427)/(1+exp(1.257*(x-2.700)))</td>
<td>0.590</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td>y = 0.103+0.443*x</td>
<td>y = 2.427+(0.017 - 2.427)/(1+exp(1.257*(x-2.700)))</td>
<td>0.590</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sum of Errors of Errors</td>
<td>0.751</td>
<td>0.126</td>
<td>0.090</td>
<td>0.090</td>
<td>0.108</td>
<td></td>
</tr>
<tr>
<td>R^2</td>
<td>0.954</td>
<td>0.992</td>
<td>0.998</td>
<td>0.995</td>
<td>0.994 (0.998*)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Plate 3</th>
<th>Model</th>
<th>Linear</th>
<th>Cubic</th>
<th>Three Parameter Logistic</th>
<th>Four Parameter Logistic with Constant Variance</th>
<th>Four Parameter Logistic with Heterogeneous Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimated</td>
<td>y = 0.003+ 0.018<em>x+ 0.279</em>x^2 - 0.037*x^3</td>
<td>y = 2.475-2.475/(1+exp(1.361*(x-2.585)))</td>
<td>y =2.448+( 0.074 - 2.448)/(1+exp(1.259*(x-2.699)))</td>
<td>1.018</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td>y = 0.095+0.453*x</td>
<td>y =2.448+( 0.074 - 2.448)/(1+exp(1.259*(x-2.699)))</td>
<td>1.018</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sum of Errors of Errors</td>
<td>1.136</td>
<td>0.163</td>
<td>0.141</td>
<td>0.124</td>
<td>0.051</td>
<td></td>
</tr>
<tr>
<td>R^2</td>
<td>0.935</td>
<td>0.991</td>
<td>0.997</td>
<td>0.993</td>
<td>0.999 (&gt;0.999*)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Plate 4</th>
<th>Model</th>
<th>Linear</th>
<th>Cubic</th>
<th>Three Parameter Logistic</th>
<th>Four Parameter Logistic with Constant Variance</th>
<th>Four Parameter Logistic with Heterogeneous Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimated</td>
<td>y = 0.009+ 0.037<em>x+ 0.249</em>x^2 - 0.03*x^3</td>
<td>y = 2.441-2.441/(1+exp(1.287*(x-2.658)))</td>
<td>y=2.422+( 0.045 - 2.422)/(1+exp(1.145*(x-2.765)))</td>
<td>1.041</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model</td>
<td>y = 0.081+0.440*x</td>
<td>y=2.422+( 0.045 - 2.422)/(1+exp(1.145*(x-2.765)))</td>
<td>1.041</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sum of Errors of Errors</td>
<td>0.903</td>
<td>0.133</td>
<td>0.115</td>
<td>0.110</td>
<td>0.066</td>
<td></td>
</tr>
<tr>
<td>R^2</td>
<td>0.945</td>
<td>0.992</td>
<td>0.997</td>
<td>0.993</td>
<td>0.995 (0.973*)</td>
<td></td>
</tr>
</tbody>
</table>
Note: *$R^2$ in the () is the weighted $R^2$
Table A5. Goodness of Fit Statistics of All Plates for Dust Mite and Cat Assays

<table>
<thead>
<tr>
<th>Number of Plates</th>
<th>Median $\omega_2$ (95% CI$^*$)</th>
<th>Median $\epsilon_2 (95% \text{ CI})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dust Mite</td>
<td>129 &gt;0.99 (0.99,1.00)</td>
<td>&gt;0.99 (0.99,1.00)</td>
</tr>
<tr>
<td>Cat</td>
<td>128 &gt;0.99 (0.97,1.00)</td>
<td>&gt;0.99 (0.97,1.00)</td>
</tr>
</tbody>
</table>

*CI is calculated by percentile method
Appendix XI. Cat Allergen Assay and Laboratory Dilutions

This paper shows that the dose-response curves of allergen assays have heterogeneous variances/precisions. An assay with good repeatability has imprecision (coefficient variation) below 20% along most part of its dose-response curve. To quantify cat allergen levels in home dust samples, a laboratory usually dilutes dust samples and obtains at least one diluted concentration level which has a good repeatability. However, the laboratory needs to dilute the home dust samples prior to knowing the cat allergen levels. In this study, the laboratory chose dilutions of 5, 25, 125, 625 for most cat dust sample assays. Many times the first assay of a home sample is not successful and none of the diluted concentrations had a contained precision. These samples have to be reanalyzed at different dilutions.

The objective of this section is to show that cat ownership can be used as a predictor of home cat allergen levels and dilutions can be adjusted according to home cat ownership status, to obtain more precise estimates of cat allergen concentrations.

In this study, 415 diluted dust samples were collected from homes with at least one cat and 920 diluted samples were from homes without a cat. These 1335 diluted samples were assayed in 131 micro titer plates. The average concentration range with good repeatability for these 131 plates is 1.78 to 44.06 µg/ml. Its corresponding OD range is 0.06 to 1.73. Each diluted home sample has a measured OD. The ODs from homes with and without a cat were tested by a two sample t test. Homes with at least one cat had significantly higher measured ODs (p<0.01). Different dilution levels are expected according to home cat ownership status. A series of simulated dilution levels are chosen from 5 to $5^{10}$ (or 9,765,625). Diluted ODs at these dilutions are simulated and shown in
Table A6. For home dust samples with cat, simulated ODs have a good repeatability (coefficient variation below 20%) at dilution level 25 and 125. For home dust samples without cat, simulated ODs have a good repeatability at dilution level 625 and 3125.

Table A6 Diluted OD by home cat ownership status.

<table>
<thead>
<tr>
<th>Simulated OD</th>
<th>Dilution</th>
<th>At Least One Cat (N=920)</th>
<th>No Cat (N=415)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>1.97</td>
<td>148.19</td>
</tr>
<tr>
<td>2</td>
<td>25</td>
<td><strong>0.39</strong></td>
<td>29.64</td>
</tr>
<tr>
<td>3</td>
<td>125</td>
<td><strong>0.08</strong></td>
<td>5.93</td>
</tr>
<tr>
<td>4</td>
<td>625</td>
<td>0.02</td>
<td><strong>1.19</strong></td>
</tr>
<tr>
<td>5</td>
<td>3125</td>
<td>&lt;0.01</td>
<td><strong>0.24</strong></td>
</tr>
<tr>
<td>6</td>
<td>15625</td>
<td>&lt;0.01</td>
<td>0.05</td>
</tr>
<tr>
<td>7</td>
<td>78125</td>
<td>&lt;0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>8</td>
<td>390625</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>9</td>
<td>1953125</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>10</td>
<td>9765625</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
</tr>
</tbody>
</table>

Therefore, the laboratory can dilute dust samples based on home cat ownership information. If a dust sample is from a home with a cat and a high OD is expected due to high cat allergen concentration, high dilution levels should be chosen to improve the precision of the estimated home allergen concentration; whereas, samples from homes without a cat that may have a lesser cat allergen concentration will require less dilution.


Ritz, C. and Streibig, J.C. Bioassay Analysis using R. http://cran.r-project.org


TableCruve2D. SYSTAT, Inc. http://www.systat.com/products/TableCurve2D

The nlmse Package. http://cran.r-project.org
