# tvst

**Methods**

# Quantifying Uncertainty of the Estimated Visual Acuity Behavioral Function With Hierarchical Bayesian Modeling

## Yukai Zhao[1], Luis Andres Lesmes[2], Michael Dorr[2], and Zhong-Lin Lu[3–5]

[1] Center for Neural Science, New York University, New York, NY, USA
[2] Adaptive Sensory Technology Inc., San Diego, CA, USA
[3] Division of Arts and Sciences, NYU Shanghai, Shanghai, China
[4] Center for Neural Science and Department of Psychology, New York University, New York, NY, USA
[5] NYU-ECNU Institute of Brain and Cognitive Science at NYU Shanghai, Shanghai, China

**Purpose:** The goal of this study is to develop a hierarchical Bayesian model (HBM) to better quantify uncertainty in visual acuity (VA) tests by incorporating the relationship between VA threshold and range across multiple individuals and tests.

**Methods:** The three-level HBM consisted of multiple two-dimensional Gaussian distributions of hyperparameters and parameters of the VA behavioral function (VABF) at the population, individual, and test levels. The model was applied to a dataset of quantitative VA (qVA) assessments of 14 eyes in 4 Bangerter foil conditions. We quantified uncertainties of the estimated VABF parameters (VA threshold and range) from the HBM and compared them with those from the qVA.

**Results:** The HBM recovered covariances between VABF parameters and provided better fits to the data than the qVA. It reduced the uncertainty of their estimates by 4.2% to 45.8%. The reduction of uncertainty, on average, resulted in 3 fewer rows needed to reach a 95% accuracy in detecting a 0.15 logMAR change of VA threshold or both parameters than the qVA.

**Conclusions:** The HBM utilized knowledge across individuals and tests in a single model and provided better quantification of the uncertainty of the estimated VABF, especially when the number of tested rows was relatively small.

**Translational Relevance:** The HBM can increase the accuracy in detecting VA changes. Further research is necessary to evaluate its potential in clinical populations.

## Introduction

Visual acuity (VA) is the most important functional vision metric in the clinic.[1] Precise assessment of VA is of paramount importance for accurate diagnosis of eye diseases, monitoring disease progression, and evaluating treatment efficacy, as well as setting classification and qualification standards in many sports and professions.[2–8] However, VA measurements are intrinsically noisy and the estimated VA scores carry a lot of uncertainty. This is because human VA behavior is probabilistic in nature.[9–12] Given the exact same optotypes, the responses from a human observer may vary from trial to trial because of the various noises in the perceptual system.[13–16] In this study, we apply the Hierarchical Bayesian Modeling (HBM) approach to improve quantification of uncertainty in VA assessments.

It is well known that human behavior in optotype identification can be modeled by the visual acuity behavioral function (VABF; Fig. 1A), that is, the probability of correct optotype identification as a function of optotype size. The function has two parameters, VA threshold, the optotype size required to reach a certain performance level, and VA range, which specifies how fast acuity behavior changes with increasing or decreasing optotype size.[17–20] Although traditional VA assessments have only focused on estimating VA threshold, both VABF parameters may vary across individuals and disease stages.[18,21–23] A complete characterization of the VABF requires a joint estimate of both parameters.[24,25]
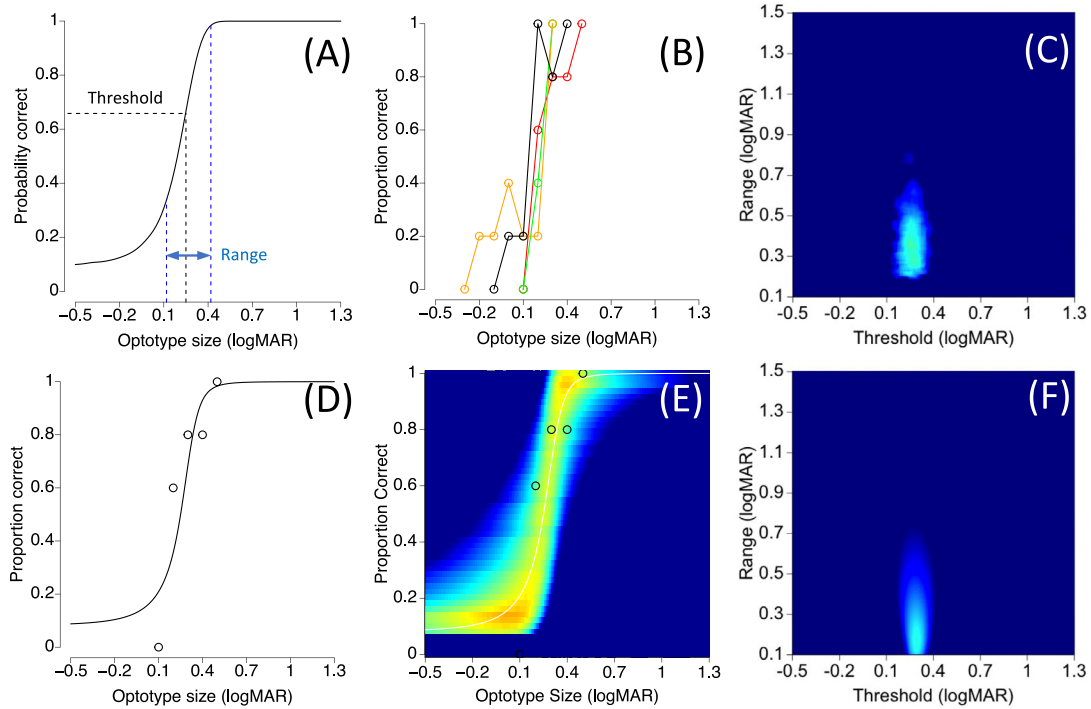
**Figure 1.** VA behavioral function (VABF) and uncertainty in VA assessments. (**A**) A VABF with two parameters: VA threshold and range. (**B**) Estimated VABFs from repeated ETDRS tests, which sample the theoretical VABF with a number of optotype sizes, each tested five times (*circles*). (**C**) A color density plot of the point VABF parameter estimates across repeated ETDRS tests. (**D**) Data (*circles*) from one ETDRS assessment and the best fitting VABF (*curve*). (**E**) Probability distribution of the estimated VABFs from **D**. (**F**) The joint probability distribution of the estimated VABF parameters from **D**.

The VABF is a theoretical construct that specifies the probability of correct optotype identification at every possible optotype size. In actual VA assessments, the function is only sampled with a finite number of optotype sizes, and each with a small number of trials. As a result, the empirical VABFs and the corresponding VABF parameters vary across repeated assessments (Figs. 1B, 1C). Figure 1C is a color density plot of the point estimates across multiple tests. It illustrates the joint probability of different point estimates of VABF parameters from many repeated measurements. In this scheme, each VA assessment results in a single estimate of the VABF parameters, so repeated assessments are required to quantify the uncertainty.

In the real world, where repeated VA assessments are most often not possible, how do we gauge uncertainty from a single VA assessment? This can be done with several modeling approaches.[26] Given the results from a single VA assessment (Fig. 1D), the probability distribution of VABF can be estimated by their likelihood or Bayesian inference (Fig. 1E). The corresponding parameters and the associated probabilities form a joint distribution of the estimated VABF parameters (Fig. 1F). The mean and spread of the joint distribution can be used to quantify the estimated parameters and their uncertainties,[26–30] which is often quantified by the half width of the 68.2% credible interval (68.2% HWCI) of the distribution,[26,27] defined as the smallest interval that contains the true value of the estimated quantity with 68.2% probability. The estimated uncertainties from repeated tests (see Fig. 1C) and a single measurement (see Fig. 1F) are equivalent under ideal experimental conditions.[26,30,31]

Traditional VA assessments, including the printed VA charts and computerized tests, sample the VABF with optotypes that are 0.1 logMAR apart, each tested on a row with a small number of optotypes.[32–34] The method of limits and heuristic termination rules are used to determine a single VA score (but see Ref. 35). Although uncertainty is usually not estimated, studies with repeated tests found large uncertainties (approximately 0.07–0.34 logMAR)[33,36–43] because of the coarse sampling of the VABF as well as the use of heuristic termination rules that depend on the probabilistic behavior of the human observer. In addition, the exact performance level (probability of correct optotype identification) associated with each heuristic termination rule depends on the range parameter of the VABF.[24,25] We cannot meaningfully compare
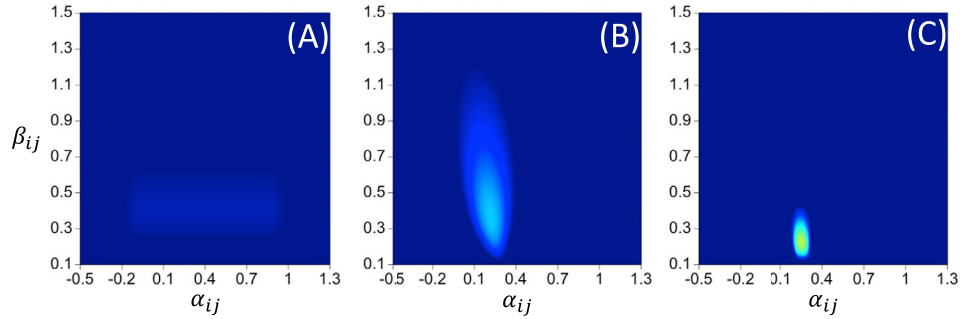
**Figure 2.** Evolution of the joint distributions of VABF parameter $\theta_{ij}$ of individual $i$ in test $j$, from prior (**A**) to posterior probability distributions after 15 (**B**), and 45 (**C**) rows in a qVA test.

VA scores from patients with different VA range parameters or interpret changes of VA scores of the same patient with range parameter changing between assessments.[18,19]

The quantitative visual acuity (qVA) method[24,25] is a Bayesian inference procedure developed to characterize the full VABF and quantify the uncertainty in each assessment of individual $i$ in test $j$. In the qVA, the VABF (see Fig. 1) is characterized with parameter $\theta_{ij} = (\alpha_{ij}, \beta_{ij})$, where $\alpha_{ij}$ is VA threshold, corresponding to the $d' = 2$ performance level, $\beta_{ij}$ is the range parameter of the function that covers $d' = 1$ to $d' = 4$ performance levels.[24,25] The qVA incorporates high density optotype size sampling and starts with a broad joint prior distribution of $\theta_{ij}$, representing existing knowledge of the general pattern of the VABF (Fig. 2A). An active learning approach is used to optimize stimulus selection to reduce the expected uncertainty of the posterior distribution of $\theta_{ij}$ in each trial (see Fig. 2). Bayes' rule is used to compute the joint posterior distribution of $\theta_{ij}$ (Figs. 2B, 2C), which allows us to quantify not only the VABF parameters but also their uncertainties from a single measurement. In computer simulations, we showed that the qVA could assess VABF parameters with virtually no bias, and very small uncertainty in $\alpha_{ij}$ (HWCI = 0.028 logMAR) and small uncertainty in $\beta_{ij}$ (HWCI = 0.092 logMAR), reflecting 52.5% and 49.5% uncertainty reductions of the estimated $\alpha_{ij}$ relative to the electronic-early treatment diabetic retinopathy study (E-ETDRS) and Freiburg Visual Acuity and Contrast Test (FrACT) methods, respectively.[25] The results were confirmed in a psychophysical study: estimated $\theta_{ij}$ from the qVA exhibited very small ($\alpha_{ij} = 0.019$ logMAR HWCI) and relatively small ($\beta_{ij} = 0.062$ logMAR HWCI) uncertainty. In addition, we found a significant correlation (r = 0.412, $P < 0.001$) between estimated $\alpha_{ij}$ and $\beta_{ij}$ across individuals and tests in the psychophysical experiment.[25]
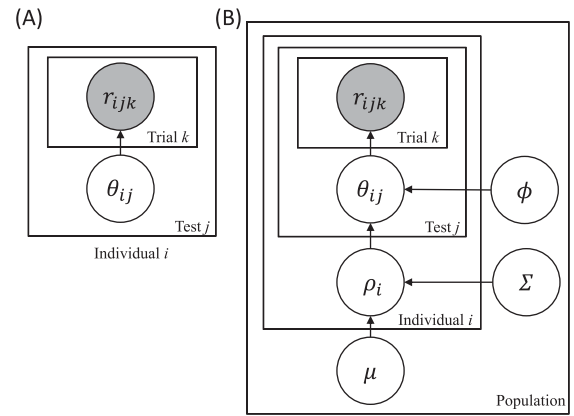


**Figure 3.** (**A**) The qVA models the VABF of individual $i$ in test $j$ with two parameters: VA threshold $\alpha_{ij}$ and range $\beta_{ij}$, $\theta_{ij} = (\alpha_{ij}, \beta_{ij})$. (**B**) A three-level hierarchical Bayesian model (HBM) of the VABF across multiple individuals and tests. At the population level, μ and Σ are the mean and covariance hyperparameters of the population. At the individual level $\rho_i$ and $\varphi$ are the mean and covariance hyperparameters of individual $i$. At the test level, $\theta_{ij}$ is the VA threshold and range of individual $i$ in test $j$. The same generative model in the qVA, the VABF ($\theta_{ij}$), determines the likelihood of response $r_{ijk}$ of individual $i$ in trial $k$ of test $j$.

In its current implementation, each qVA test (Fig. 3A) starts with a broad prior. It does not take advantage of any knowledge across individuals and tests such as the observed correlation between $\alpha_{ij}$ and $\beta_{ij}$,[24,25] which could be useful in further constraining the estimated $\theta_{ij}$s and thereby reducing their uncertainties. In other words, the broad prior and independent treatment of each qVA test may overestimate the uncertainty of the estimated VABF. Although the qVA has greatly reduced the uncertainty of $\alpha_{ij}$ estimates, the uncertainty of $\beta_{ij}$ estimates is still relatively large.[25]

The observed correlation between estimated $\alpha_{ij}$ and $\beta_{ij}$ across individuals and tests in our previous study[25] motivated us to develop a three-level HBM (Fig. 3B) to quantify uncertainties at the population,

individual, and test levels in a single model. At the population level, $p(\eta)$, the joint distribution of VABF hyperparameter $\eta$ across all the individuals, is modeled as a mixture of 2-dimensional Gaussian distributions with mean μ and covariance Σ, which have distributions $p(\mu)$ and $p(\Sigma)$. At the individual level, $p(\tau_i|\eta)$, the joint distribution of VABF hyperparameter $\tau_i$ of individual $i$ across all the tests performed by the individual, is modeled as a mixture of 2-dimensional Gaussian distributions with mean $\rho_i$ and covariance $\varphi$, which have distributions $p(\rho_i|\eta)$ and $p(\varphi)$. $p(\rho_i|\eta)$ denotes that $\rho_i$ is conditioned on $\eta$. At the test level, $p(\theta_{ij}|\tau_i)$, the joint distribution of the VABF parameters of individual $i$ in test $j$, $\theta_{ij}$, is conditioned on $\tau_i$. The cross- and within-individual regularities are modeled as covariances Σ and $\varphi$ in the HBM.[44–48] Bayes' rule is then used to update the joint prior distribution of all the parameters and hyperparameters to their joint posterior distribution,[49–53] which allows us to quantify the estimated hyperparameters and parameters as well as their uncertainties at the population, individual and test levels.[49,54,55]

The HBM has been used in many different disciplines to model data with hierarchical structures, including astronomy,[56] ecology,[57,58] genetics,[59] and cognitive science.[50,51,53,60–65] By decomposing the variability of an entire dataset into distributions at multiple levels of the hierarchy, it can better quantify uncertainty at each level.[59,52,65] Here, we developed an HBM and evaluated its performance relative to that of the qVA using the qVA data in 14 eyes tested in 4 Bangerter foil conditions.[25] Our goal was to improve quantification of uncertainty at the test level. We hypothesized that the HBM would account for the data better than the qVA and reduce the uncertainties of estimated VABF parameters in single VA tests relative to the qVA.

# Methods

## Data

The dataset used in this study included a total of 56 tests: 14 eyes (left and right eyes of 7 subjects), each tested with the qVA with 3 levels of Bangerter foils[66,67] and without foil.[25] Each qVA test consisted of 45 rows ($K = 45$) of 3 optotypes (Fig. 4). In each qVA trial, 3 letters on a row, randomly sampled without replacement from the 10 Sloan letters (C, D, H, K, N, O, R, S, V, and Z) and with their size determined by the qVA, were presented to the subject. The subject was instructed to report the identity of the letters verbally. The joint posterior distributions of VABF parameters



| Row # | | logMAR |
|---|---|---|
| 1 | R D C | 0.40 |
| 2 | C R N | 0.12 |
| 3 | K R D | -0.08 |
| | ⋮ | |
| 43 | N O K | 0.04 |
| 44 | Z R K | 0.08 |
| 45 | Z H C | -0.18 |

**Figure 4.** A sample stimulus sequence in a qVA test. Stimuli on rows 1, 2, 3, 43, 44, and 45 are shown, with three optotypes in each row.

in each qVA test after $k = 5, 10, 15, 20, 25, 30, 35, 40,$ and 45 rows were used to conduct statistical analysis. The details of the qVA can be found in prior publications.[24,25]

## Apparatus

All analysis was conducted on a Dell computer with Intel Xeon W-2145 @ 3.70 GHz CPU (8 cores and 16 threads) and 64 GB installed memory (RAM). The qVA was implemented in Matlab R2013a (MathWorks Corp., Natick, MA, USA) and the HBM was implemented with JAGS[68] in R.[69]

## The HBM

### Likelihood Function for a Single Test

The likelihood function of a model is defined as the probability of the observed data given a particular set of parameter values in the model. The VABF (see Fig. 1A) in qVA was used to compute the probability of individual $i$'s response $r_{ijk}$ (the number of correct identifications: 0, 1, 2, or 3) of the 3 optotypes with optotype size $os_{ijk}$ in row $k$ of test $j$:

$$p\left(r_{ijk}|\theta_{ij}, os_{ijk}\right) = f\left(g\left(os_{ijk}, \theta_{ij}\right), n\right), \quad (1a)$$

where $g(os_{ijk}, \theta_{ij})$ is the VABF of correct identification of a single optotype with size $os_{ijk}$ and model parameters $\theta_{ij} = (\alpha_{ij}, \beta_{ij})$; $f(g(os_{ijk}, \theta_{ij}), n)$ is the probability of observing the number of correct identifications given $g(os_{ijk}, \theta_{ij})$ and a specific chart design (e.g. $n = 3$ optotypes on a row). The details of functions $f$ and $g$ can be found in Supplementary Materials A and previous publications.[24,43] The probability of obtaining the observed responses in all rows in test $j$ from individual $i$ is the product of the probability of responses in all

rows (Equation 1a):

$$p\left(r_{ij,1:K}|\theta_{ij}, os_{ij,1:K}\right) = \prod_{k=1}^{K} p\left(r_{ijk}|\theta_{ij}, os_{ijk}\right). \quad (1b)$$

### Three-Level Hierarchy

To account for the entire dataset with all the individuals and tests, we constructed an HBM with three levels (see Fig. 3). In this model, the population level distribution of VABF hyperparameters sets the prior for the mean hyperparameters of each individual, which in turn sets the prior of the parameters of the individual in all her/his tests. We describe the distributions at these three levels in more details below.

At the population level, $p(\eta)$, the joint distribution of hyperparameter $\eta$, is modeled as a mixture of two-dimensional Gaussian distributions $\mathcal{N}$ with mean $\mu$ and covariance $\Sigma$, which have distributions $p(\mu)$ and $p(\Sigma)$:

$$p\left(\eta\right) = \mathcal{N}\left(\eta, \mu, \Sigma\right) p\left(\mu\right) p\left(\Sigma\right). \quad (2)$$

At the individual level, $p(\tau_i|\eta)$, the joint distribution of hyperparameter $\tau_i$ of individual $i$, is modeled as a mixture of two-dimensional Gaussian distributions $\mathcal{N}$ with mean $\rho_i$ and covariance $\varphi$, which have distributions $p(\rho_i|\eta)$ and $p(\varphi)$:

$$p\left(\tau_i|\eta\right) = \mathcal{N}\left(\tau_i, \rho_i, \phi\right) p\left(\rho_i|\eta\right) p\left(\phi\right), \quad (3)$$

where $p(\rho_i|\eta)$ denotes that $\rho_i$ is conditioned on $\eta$.

At the test level, $p(\theta_{ij}|\tau_i)$, the joint distribution of parameter $\theta_{ij}$ is conditioned on $\tau_i$.

The probability of obtaining the entire dataset is computed by probability multiplication:

$$p\left(r_{1:I,1:J,1:K}|X, os_{1:I,1:J,1:K}\right) =$$
$$\prod_{i=1}^{I} \prod_{j=1}^{J} p\left(r_{ij,1:K}|\theta_{ij}, os_{ij,1:K}\right) p\left(\theta_{ij}|\tau_i\right) p\left(\tau_i|\eta\right) p\left(\eta\right)$$
$$= \prod_{i=1}^{I} \prod_{j=1}^{J} p\left(r_{ij,1:K}|\theta_{ij}, os_{ij,1:K}\right) p\left(\theta_{ij}|\tau_i\right) \mathcal{N}\left(\tau_i, \rho_i, \phi\right)$$
$$p\left(\rho_i|\eta\right) p\left(\phi\right) \mathcal{N}\left(\eta, \mu, \Sigma\right) p\left(\mu\right) p\left(\Sigma\right), \quad (4)$$

where $X = (\theta_{1:I, 1:J}, \rho_{1:I}, \varphi, \mu, \Sigma)$ are all the parameters and hyperparameters in the HBM, $I$ is the total number of individuals, and $J$ is the total number of tests on each individual.

### Bayesian Inference

We start with prior distributions of $\mu$, $\Sigma$, and $\varphi$.

$$p_0\left(\mu\right) = \mathcal{U}\left(\mu_{0,min}, \mu_{0,max}\right), \quad (5a)$$

$$p_0\left(\Sigma^{-1}\right) = \mathcal{W}\left(\Sigma_{qVA}^{-1}/\nu, \nu\right), \quad (5b)$$

$$p_0\left(\phi^{-1}\right) = \mathcal{W}\left(\phi_{qVA}^{-1}/\nu, \nu\right), \quad (5c)$$

where $\mathcal{U}$ is a two-dimensional uniform distribution with $\mu_{0,min} = (-0.5, 0.1)$ logMAR and $\mu_{0,max} = (1.3, 1.5)$ logMAR; precision matrices $\varphi^{-1}$ and $\Sigma^{-1}$ are the inverse of covariance $\varphi$ and $\Sigma$; $\mathcal{W}(\frac{Y}{\nu}, \nu)$ denotes a Wishart distribution with expected mean precision matrix Y and degrees of freedom $\nu = 2$; $\phi_{qVA}^{-1}$ is the inverse of the average covariance matrix $\varphi_{qVA}$ computed from the joint posterior distributions across all qVA tests in the dataset; $\Sigma_{qVA}^{-1}$, is the inverse of the covariance matrix $\Sigma_{qVA}$ computed from the estimated $\theta_{ij}$ across all qVA tests.

Bayes' rule is used to compute the joint posterior distribution of all the parameters and hyperparameters in the HBM:

$$p\left(X|r_{1:I,1:J,1:K}, os_{1:I,1:J,1:K}\right) =$$
$$\frac{\prod_{i=1}^{I} \prod_{j=1}^{J} p(r_{ij,1:K}|\theta_{ij}, os_{ij,1:K})p(\theta_{ij}|\tau_i)\mathcal{N}(\tau_i, \rho_i, \phi)p(\rho_i|\eta)p_0(\phi)\mathcal{N}(\eta, \mu, \Sigma)p_0(\mu)p_0(\Sigma)}{\int \prod_{i=1}^{I} \prod_{j=1}^{J} p(r_{ij,1:K}|\theta_{ij}, os_{ij,1:K})p(\theta_{ij}|\tau_i)\mathcal{N}(\tau_i, \rho_i, \phi)p(\rho_i|\eta)p_0(\phi)\mathcal{N}(\eta, \mu, \Sigma)p_0(\mu)p_0(\Sigma)dX}, \quad (6)$$

where the denominator is the integral of the probability of obtaining the entire dataset (Equation 4) across all possible values of $X$ and is a constant for a given dataset and HBM.

### Computing the Joint Posterior Distribution

In this study, an individual refers to a combination of subject, eye and foil level, with the total number of individuals $I = 56$. In addition, each individual was only tested once ($J = 1$).

The complexity of the HBM increases exponentially with the number of hyperparameters and parameters. It would take practically infinite time to compute all the points in the 232-dimensional hyperparameter/parameter space defined in Equation 6. We used the JAGS[68] package in R[69] to evaluate the joint posterior distribution (see Equation 6). JAGS uses an efficient sampling algorithm to generate representative samples of the joint posterior distribution of all the parameters and hyperparameters in the HBM via Markov Chain Monte Carlo (MCMC). Each MCMC generated 15,000 samples via a random walk process (see Supplementary Materials B for details). Because the initial part of the random walk was largely influenced by the arbitrary starting position, steps in the burn-in and adaptation (further optimization of the sampling algorithm by JAGS) phases were discarded and not included in the analysis. The exact number of steps discarded differ for each model. In this study, 20,000 and 100,000 steps were used for burn-in and adaptation in each HBM model based on results from pilot studies. Convergence of each parameter was evaluated with Gelman and Rubin's diagnostic rule[70] based on the

ratio of between- and within-MCMC variances along each dimension, that is, the variance of the samples across MCMC processes divided by the variance of the samples in each MCMC process. The HBM was deemed "converged" when the ratios for all the parameters were smaller than 1.05. The data with $k = 5, 10, 15, 20, 25, 30, 35, 40,$ and 45 rows were fit by the HBM separately.

## Simulations

Each simulated dataset consisted of 56 qVA tests from 56 individuals, each tested once. The VABF parameters for the simulated observers were randomly sampled from the posterior distributions of $\theta_{i1}$ obtained from the HBM fit to the real data. Each qVA test consisted of 45 three-optotype rows, and the row-by-row responses were determined by the VABF specified with the parameters of the simulated observer. The HBM was fit to the simulated dataset of 56 tests. The procedure was repeated 10 times.

## Statistical Analysis

### Goodness-of-Fit
Bayesian predictive information criterion (BPIC)[71,72] was used to quantify and compare the goodness-of-fit of the HBM against that of the qVA. The BPIC quantifies the likelihood of the data based on the joint posterior distribution of the parameters of a model and penalizes model complexity.

### Uncertainty
We compared the uncertainties of estimated $\theta_{i1}$ from the qVA and HBM. The uncertainty of each estimated parameter was quantified by the 68.2% HWCI of its marginal posterior distribution.[26,27] Paired *t*-test (R function *t.test*[69]) was used to compare the 68.2% HWCIs between the two methods.

### Discrimination Accuracy
To demonstrate effects of uncertainty reduction, we computed the accuracy in detecting a 0.15 logMAR VA threshold ($\alpha_{i1}$) change, a 0.15 logMAR range ($\beta_{i1}$) change, and a 0.15 logMAR change of both parameters. We chose 0.15 logMAR as the magnitude of $\alpha_{i1}$ change because a greater-than-15-letter VA improvement (0.3 logMAR) is considered by the US Food & Drug Administration (FDA) as an acceptable end point of a clinical trial, although a greater-than-10-letter VA improvement (0.2 logMAR) has been used when the benefits can outweigh the safety risks of the proposed method or product.[73] We chose 0.15 logMAR as the magnitude of $\beta_{ij}$ change, because a

similar change has been reported in eyes with degraded vision.[18,25]

Discrimination accuracy was quantified by the area under receiver operating characteristic curve (AUROC). Given a change-criterion, specificity is defined as the probability of correctly identifying no change, whereas sensitivity is defined as the probability of correctly identifying a change.[74] Whereas sensitivity and specificity depend on the change criterion, the AUROC in signal detection theory provides a criterion-free estimate of the accuracy of a test to detect a change.[74] It only depends on the discriminability $d'$ that weighs the evidence against uncertainty and quantifies the signal (change) to noise (HWCI) ratio of two probability distributions of the estimates in two conditions.

For one-dimensional distributions, $d'$ is defined as[74]:

$$d' = \frac{\Delta}{\mathrm{HWCI}}, \qquad (7)$$

where $\Delta$ is the magnitude of change. For multidimensional difference distributions, $d'$ is defined as[75]:

$$d' = \sqrt{\Delta * cov(\Delta)^{-1} * \Delta^{T}}, \qquad (8a)$$

$$cov(\Delta) = \begin{bmatrix} \mathrm{HWCI}^2_{threshold} & covariance \\ covariance & \mathrm{HWCI}^2_{range} \end{bmatrix}, \qquad (8b)$$

$$covariance = corr \times \mathrm{HWCI}_{threshold} \times \mathrm{HWCI}_{threshold}$$

where $\Delta = (0.15, 0.15)$, $cov(\Delta)$ is the covariance matrix of the VABF parameter difference distribution, $cov(\Delta)^{-1}$ is the inverse of $cov(\Delta)$, $\Delta^{T}$ is the transpose of $\Delta$, * represents matrix multiplication, and *corr* is the correlation coefficient at the test level in the HBM.

## Results

### Goodness-of-Fit

The BPIC for the qVA and HBM were 4867 and 4832, respectively, indicating that the HBM fit the data better than the qVA. Figure 5 shows the posterior distributions of the estimated VABFs in one test from the qVA and HBM after 5, 15, and 45 rows. Although the HBM fit the data better and reduced the uncertainty of the estimated $\theta_{1:I,1}$, it did not significantly change the expected values of the estimated $\theta_{1:I,1}$: the average absolute differences of the estimates from the qVA and HBM across all individuals were 0.002 and 0.022 logMAR, respectively, indicating that HBM and qVA estimates exhibited virtually the same central tendency.
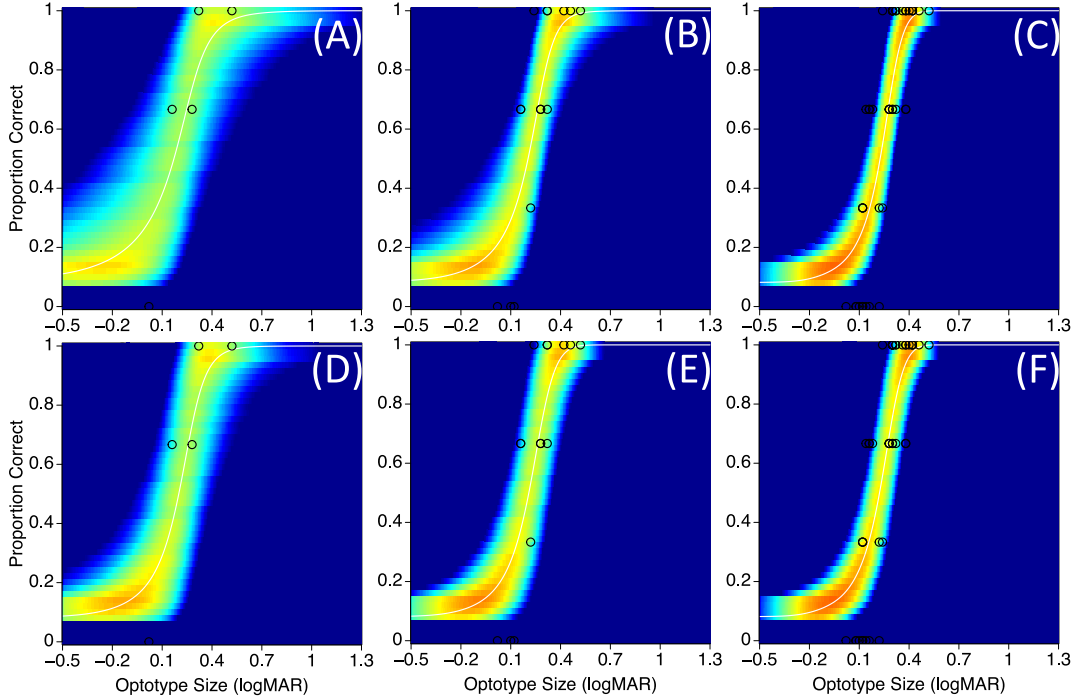
**Figure 5.** Posterior distributions of the estimated VABFs in one test from the qVA (**A, B, C**) and HBM (**D, E, F**) after 5, 15, and 45 rows. *Open circles* represent data.

## Posterior Distributions From the HBM

Figure 6 illustrates the joint posterior distributions of hyperparameters μ, Σ, and η after 45 rows. Whereas μ and Σ were from the HBM fits to the data (see Equation 6), η was constructed from the distributions of μ and Σ (see Equation 2). The correlation coefficient between $\eta^1$ and $\eta^2$ was 0.662. ($P < 0.001$). Figure 7 illustrates the joint posterior distributions of hyperparameters $\varphi$, and $\rho_i$, and $\tau_i$ for 8 out of the 56 individuals after 45 rows. Whereas $\rho_i$ and $\varphi$ were from the HBM fits to the data (see Equation 6), $\tau_i$ was constructed from the distributions of $\rho_i$ and $\varphi$ (see

Equation 3). Figure 8A illustrates the joint posterior distributions of parameters $\theta_{i1}$ from the 8 corresponding individuals in Figure 7 after 45 rows. Figure 8B shows the mean $\theta_{i1}$ of all the 56 tests.

## Uncertainty of $\theta_{i1}$ Estimates

Figure 9 shows the estimated joint posterior distributions of $\theta_{i1}$ after 5, 15, and 45 rows for one individual, obtained from both the qVA and HBM. Across all the individuals and tests, the average 68.2% HWCI of the estimated $\theta_{i1}$ decreased with the number of rows in both methods (Fig. 10, Table 1). With 5 rows, the
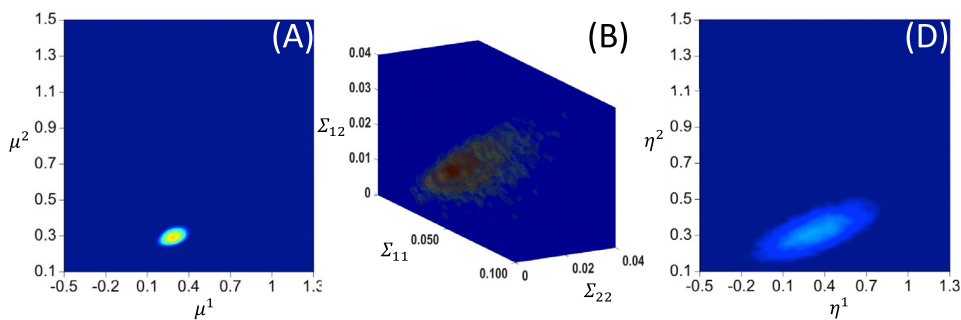


**Figure 6.** Illustrations of the joint posterior distributions of hyperparameters μ (**A**), Σ (**B**), and η (**C**) after 45 rows at the population level in the HBM.
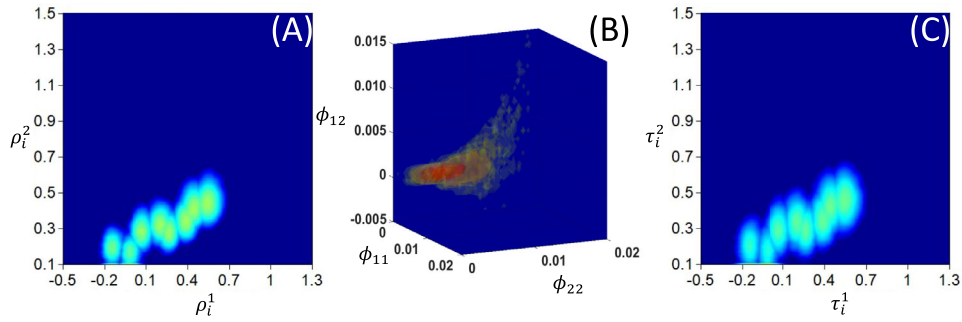
**Figure 7.** Illustrations of the joint posterior distributions of hyperparameters $\rho_i$ (**A**), $\varphi$ (**B**), and $\tau_i$ (**C**) for 8 out of the 56 individuals after 45 rows in the HBM.
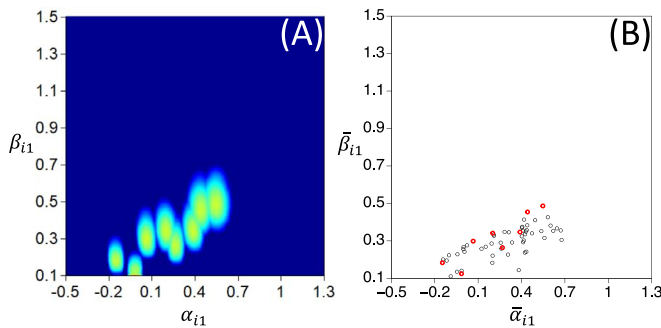


**Figure 8.** (**A**) Illustration of the joint posterior distributions of parameters $\theta_{i1}$ from 8 tests of the 8 individuals in Figure 7 after 45 rows in the HBM. (**B**) Mean $\theta_{i1}$, $\bar{\theta}_{i1} = (\bar{\alpha}_{i1}, \bar{\beta}_{i1})$, (*circles*) of all the 56 tests, with the 8 tests in **A** highlighted in *red*.

68.2% HWCIs of the estimated $\alpha_{i1}$ (t(55) = 14.8, $P < 0.001$) and $\beta_{i1}$ (t(55) = 13.7, $P < 0.001$) were significantly different between the methods. With 45 rows, the 68.2% HWCIs of the estimated $\alpha_{i1}$ (t(55) = 4.63, $P < 0.001$) and $\beta_{i1}$ (t(55) = 6.74, $P < 0.001$) were significantly different between the methods. Relative to the qVA, the HBM reduced the HWCI of the estimated $\alpha_{i1}$ by 4.2% to 26.9%, and $\beta_{i1}$ by 20.8% to 45.8%.

Uncertainty of the estimated parameters is quantified as the HWCI of the (marginal) posterior distribution of $\theta_{i1}$. Because the trial-by-trial data and distribution of $\tau_i$ were different for each individual and test, the uncertainties were different across individuals and tests in the HBM (Fig. 11).

## Simulations

The HBM recovered the parameters of the simulated tests very well. The HBM solutions had near-zero bias for $\alpha_{i1}$ ($-0.0024$ logMAR) and $\beta_{i1}$ (0.0057 logMAR) across all the simulations. The average root–mean-square errors (RMSEs) between the HBM estimates and the true values (0.019 and

0.047 logMAR for $\alpha_{i1}$ and $\beta_{i1}$) were comparable to the HWCIs of the parameter estimates (0.019 and 0.050 logMAR for $\alpha_{i1}$ and $\beta_{i1}$), indicating good convergence of the HBM fits. In addition, the HWCIs from the simulations were also comparable to those from the experimental data (0.018 and 0.048 logMAR for $\alpha_{i1}$ and $\beta_{i1}$), suggesting that the simulations captured the uncertainties in the real experiment.

## Discrimination Accuracy

Discrimination accuracy increased with the number of rows (Table 2, Fig. 12). The HBM estimates exhibited higher accuracy than the qVA, especially when the number of rows was relatively small. To reach 95% accuracy, 10 and 7 rows were needed to detect a 0.15 logMAR $\alpha_{i1}$ change with the qVA and HBM, and 10 and 6 rows were needed to detect a 0.15 logMAR change of both $\alpha_{i1}$ and $\beta_{i1}$, respectively.

## Discussion

In this study, we developed an HBM to improve uncertainty quantification in VA assessment. By explicitly quantifying regularities at the population, individual, and test levels, the HBM utilized knowledge of VABF across multiple assessments to constrain parameter estimates, and decompose the variability of an entire dataset into distributions at three levels. Applying the model to a dataset that consisted of qVA assessments of 14 eyes in 4 Bangerter foil conditions, we showed that the HBM provided better fits to the data than the qVA and reduced the HWCI of the $\alpha_{i1}$ estimates by 26.9%, 12.2%, and 4.2%, and $\beta_{i1}$ estimates by 45.8%, 35.3%, and 20.8% for tests with 5, 15, and 45 rows, respectively. Simulations showed that the HBM recovered $\alpha_{i1}$ spanning about 0.8 logMAR ($-0.15$ to
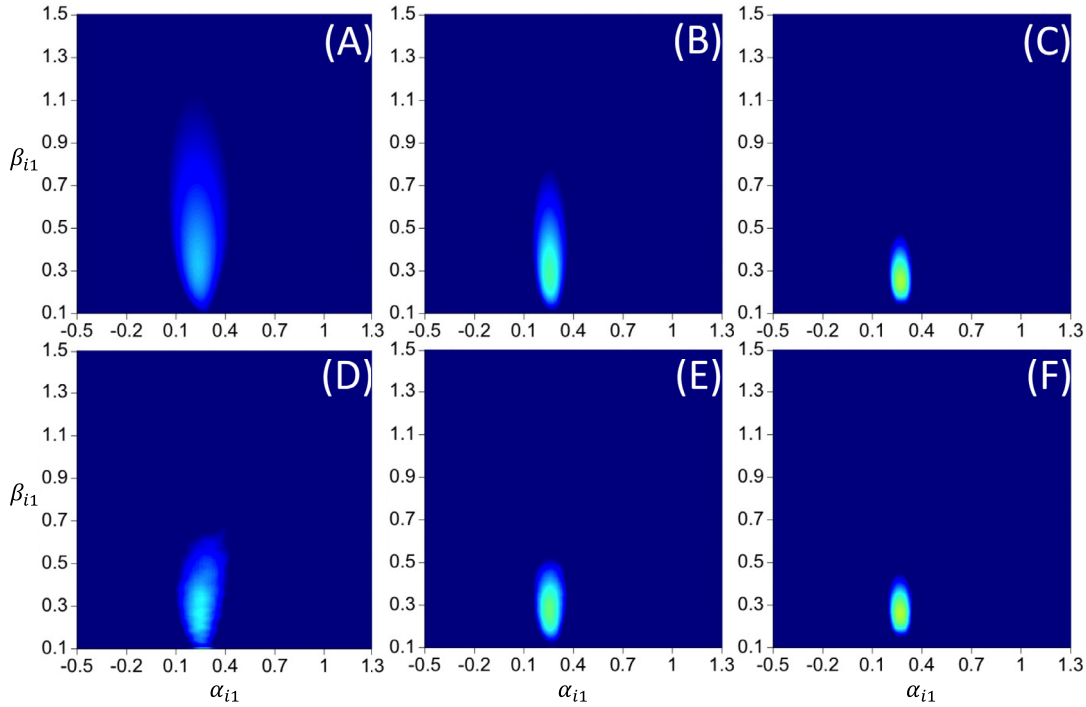
**Figure 9.** Joint posterior distributions of $\theta_{i1}$ for one individual in one qVA test after 5 (**A**), 15 (**B**), and 45 (**C**) rows. Joint posterior distributions of the corresponding $\theta_{i1}$ of the same test in the HBM after 5 (**D**), 15 (**E**), and 45 (**F**) rows.
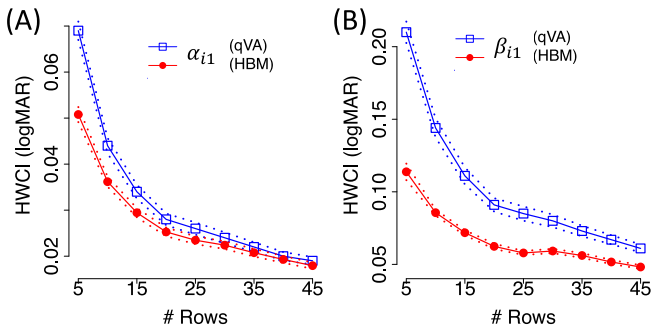


**Figure 10.** Average 68.2% HWCI ($\pm$ standard error: dotted lines) of the estimated $\alpha_{i1}$ (**A**) and $\beta_{i1}$ (**B**) as functions of number of tested rows in the qVA (*open squares*) and HBM (*filled circles*).

approximately 0.68 logMAR) and $\beta_{i1}$ spanning about 0.4 logMAR (approximately 0.11 to 0.49 logMAR). The simulation results also provided validity for the discrimination accuracy analysis.

The HBM exhibited its largest advantage relative to the qVA when the number of tested rows in each qVA test was small (see Table 1). This is because the HBM reduces uncertainty by incorporating knowledge across all individuals and tests. As the number of rows increased, more knowledge in each qVA test became available and provided better constraints on the parameters. The results highlight the importance of the HBM when fewer rows are tested, which is often the

case in clinical settings. In fact, the current HBM can be readily used to analyze data from other VA tests, such as the ETDRS (Lu et al. IOVS. 2021;62: ARVO Abstract 3546329).

The HBM reduced the uncertainty of the estimated $\theta_{i1}$ and improved the discrimination accuracy of the tests. The reduction of uncertainty, on average, resulted in 3 fewer rows needed to reach 95% accuracy in detecting changes of 0.15 logMAR in $\alpha_{i1}$ or both $\alpha_{i1}$ and $\beta_{i1}$ than the qVA (see Table 2, Fig. 12). The increased accuracy in detecting $\beta_{i1}$ changes in the HBM may enable a new range endpoint in functional vision,[5] because it has been documented that poor VA thresholds were accompanied by wider VA ranges.[25] In addition, using the joint distribution of $\theta_{i1}$ can further increase the accuracy in detecting changes of VA behavior because the two-dimensional distributions may contain more information than the one-dimensional marginal distributions.[76–78]

In this paper, we constructed a simple three-level HBM without considering any structure related to test conditions because our goal was to quantify the uncertainty of the estimated VABF in each test. The HBM can be extended to model regularities between different experimental conditions (e.g. before and after treatment) with covariance. A positive covariance between conditions can enhance the ability to detect differences between conditions, and therefore improve

**Table 1.** Average 68.2% HWCI (logMAR) of the Estimated $\alpha_{i1}$ and $\beta_{i1}$ in the qVA and HBM

| | | Number of Rows | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 5 | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 |
| $\alpha_{i1}$ | qVA | 0.069 | 0.044 | 0.034 | 0.028 | 0.026 | 0.024 | 0.022 | 0.020 | 0.019 |
| | HBM | 0.051 | 0.036 | 0.029 | 0.025 | 0.023 | 0.022 | 0.021 | 0.019 | 0.018 |
| | Reduction | 26.9% | 17.3% | 12.2% | 9.2% | 8.4% | 6.3% | 5.6% | 4.9% | 4.2% |
| $\beta_{i1}$ | qVA | 0.210 | 0.144 | 0.111 | 0.091 | 0.085 | 0.080 | 0.073 | 0.067 | 0.061 |
| | HBM | 0.114 | 0.086 | 0.072 | 0.062 | 0.058 | 0.059 | 0.056 | 0.052 | 0.048 |
| | Reduction | 45.8% | 40.5% | 35.3% | 31.1% | 31.7% | 26.2% | 23.3% | 22.8% | 20.8% |
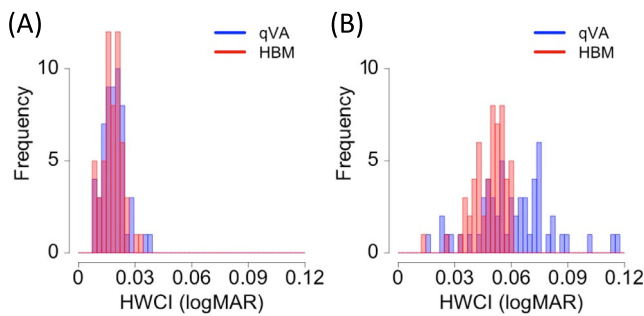


**Figure 11.** Histograms of 68.2% HWCI of the estimated $\alpha_{i1}$ (**A**) and $\beta_{i1}$ (**B**) for tests in the qVA (*blue*) and HBM (*red*) after 45 rows.

**Table 2.** Discrimination Accuracy (%) in Detecting a 0.15 logMAR $\alpha_{i1}$ Change, a 0.15 logMAR $\beta_{i1}$ Change, and a 0.15 logMAR Change of Both Parameters

| | | Number of Rows | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 5 | 10 | 15 | 20 | 25 | 30 | 35 | 40 | 45 |
| $\alpha_{i1}$ | qVA | 86.4 | 96.0 | 98.9 | 99.7 | 99.9 | 100.0 | 100.0 | 100.0 | 100.0 |
| | HBM | 93.4 | 98.3 | 99.6 | 99.9 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| $\beta_{i1}$ | qVA | 64.4 | 70.4 | 75.5 | 80.1 | 81.7 | 83.0 | 85.2 | 87.3 | 89.6 |
| | HBM | 75.0 | 81.4 | 85.6 | 89.0 | 90.7 | 90.2 | 91.4 | 93.1 | 94.4 |
| $\theta_{i1}$ | qVA | 86.8 | 96.2 | 99.0 | 99.8 | 99.9 | 100.0 | 100.0 | 100.0 | 100.0 |
| | HBM | 94.2 | 98.6 | 99.7 | 99.9 | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |

statistical power in clinical trials (unpublished communication: Zhao Y, Lesmes LA, Hou F, Lu Z-L. Hierarchical Bayesian modeling of contrast sensitivity functions in a within-subject design, 2021). The HBM can also be extended to model repeated tests of the same individual, with parameters $\theta_{ij}$ of all the tests conditioned on $\tau_i$.

Without explicit consideration of the full experimental design (e.g. two eyes from each subject), the current HBM may under- or overestimate the uncertainty from the tests, depending on the nature of the correlation (positive or negative) and experiment design. To check the impact of the simplification, we ran 50 iterations of the HBM on the data from one randomly selected eye of the 7 subjects in 4 conditions.

The procedure eliminated the correlation between eyes. We obtained virtually the same estimated $\theta_{i1}$ as the original HBM fit to the 14 eyes, with slightly higher HWCIs of the estimated $\beta_{i,1}$ (0.018 ± 0.00082 and 0.050 ± 0.00184 logMAR for $\alpha_{i1}$ and $\beta_{i1}$ in the 7 eye fits, versus 0.018 and 0.048 logMAR in the 14 eye fit). The RMSEs between the HBM fits to the 7 and 14 eyes were 0.00078 and 0.00793 logMAR for $\alpha_{i1}$ and $\beta_{i1}$, and 0.00036 and 0.00350 logMAR for the corresponding HWCIs. The results also suggest that the original experimental dataset with 14 eyes tested in 4 conditions was sufficient to constrain the HBM because cutting the amount of data to half generated virtually the same estimates.

Uncertainty in assessing the probabilistic VA behavior is determined by three factors: prior knowledge of the VA behavior, data collected in the assessment, and data analysis method. We developed the HBM as a data analysis tool in this study. The posterior distributions of the hyperparameters $\eta$ and $\tau_i$ at the population and individual levels from the HBM can also be used to construct informative priors (see Figs. 6C, 7C) within the hierarchical adaptive design optimization (HADO) framework[79,80] for new individuals (see Fig. 6C) and repeated tests of the same individual (see Fig. 7C), respectively. In the HADO framework, the joint distribution in the HBM is updated after each test and used as an informative prior in the next test to improve test efficiency for new individuals or repeated tests of the same individual. The 6-hour computation time of the HBM on a desktop computer makes it realistic to incorporate it within the HADO framework with daily updates of the joint posterior distribution.

## Conclusion

We developed a three-level HBM to utilize knowledge of VA behavior across multiple individuals and tests by explicitly modeling cross- and within-individual covariances. The HBM reduced the
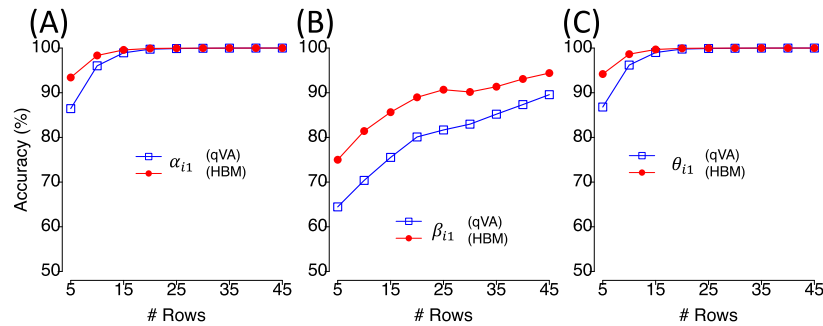
**Figure 12.** Discrimination accuracy in detecting a 0.15 logMAR $\alpha_{i1}$ change (**A**), a 0.15 logMAR $\beta_{i1}$ change (**B**), and a 0.15 logMAR change of both parameters (**C**) as functions of number of tested rows in the qVA (*open squares*) and HBM (*filled circles*).

uncertainties of the estimated VABF parameters, especially when the number of tested rows was relatively small. We applied the HBM to a qVA dataset of normal healthy subjects with simulated visual degradations. The benefit of the HBM in real clinical populations needs further research.

# Acknowledgments

Disclosure: **Y. Zhao**, None; **L.A. Lesmes**, (I, E, P); **M. Dorr**, (I, E, P); **Z.-L. Lu**, (I, P)

# References

1. Kniestedt C, Stamper RL. Visual acuity and its measurement. *Ophthalmol Clin North Am.* 2003;16:155–170.
2. Clare G, Pitts JA, Edgington K, Allan BD. From beach lifeguard to astronaut: Occupational vision standards and the implications of refractive surgery. *Br J Ophthalmol.* 2010;94:400–405.
3. World Para Alpine Skiing Classification Rules and Regulations, August 2017. Available at: https://www.paralympic.org/sites/default/files/document/170803125229547_World%2BPara%2BAlpine%2BSkiing%2BClassification%2BRules%2Band%2BRegulations_Final.pdf.
4. Kiel AW, Butler T, Alwitry A. Visual acuity and legal visual requirement to drive a passenger vehicle. *Eye.* 2003;17(5): 579–582.
5. Levenson JH, Kozarsky A. Visual Acuity. In: Walker HK, Hall WD, Hurst JW, eds. *Clinical Methods: The History, Physical, and Laboratory Examinations* (3rd ed.). Boston, US: Butterworths; 1990.
6. Rahi JS, Cumberland PM, Peckham C.S. Does amblyopia affect educational, health, and social outcomes? Findings from 1958 British birth cohort. *Br Med J.* 2006;332(7545):820–824.
7. National Research Council (US) Committee on Disability Determination for Individuals with Visual Impairments; Lennie P, Van Hemel SB, editors. Washington, DC: National Academies Press (US); 2002.
8. National Institute for Health and Care Excellence. UK National Institute for Health and Care Excellence. Age-related macular degeneration NICE guideline (NG82). January 2018. Available at: https://www.nice.org.uk/guidance/ng82.
9. Bailey I, Bullimore M, Raasch T, Taylor H. Clinical grading and the effects of scaling. *Invest Ophthalmol Vis Sci.* 1991;32:422–432.
10. Carkeet A. Modeling logMAR visual acuity scores: Effects of termination rules and alternative forced-choice options. *Optom Vis Sci.* 2001;78:529–538.
11. Loviekitchin J. Validity and reliability of visual-acuity measurements. *Ophthalmic Physiol Opt.* 1988;8:363–370.
12. Raasch TW, Bailey IL, Bullimore MA. Repeatability of visual acuity measurement. *Optom Vis Sci.* 1998;75:342–348.
13. Barlow H. Retinal noise and absolute threshold. *J Optical Society of Am.* 1956;46:634–639.
14. Faisal AA, Selen LP J, Wolpert DM. Noise in the nervous system. *Nat Rev Neurosci.* 2008;9:292–303.
15. Lu Z-L, Dosher BA. Characterizing observers using external noise and observer models:

Assessing internal representations with external noise. *Psychol. Rev.* 2008;115:44–82.

16. Neri P. How inherently noisy is human sensory processing? *Psychon Bull Rev.* 2010;17: 802–808.

17. Alexander R, Xie W, Derlacki DJ. Visual acuity and contrast sensitivity for individual Sloan letters. *Vision Res.* 1997;37:813–819.

18. Carkeet A, Lee L, Kerr JR, Keung MM. The slope of the psychometric function for Bailey-Lovie letter charts: Defocus effects and implications for modeling letter-by-letter scores. *Optom Vis Sci.* 2001;78(2):113–121.

19. Horner DG, Paul AD, Katz B, Bedell HE. Variations in the slope of the psychometric acuity function with acuity threshold and scale. *J Optom.* 1985;62:895–900.

20. Strasburger H. Converting between measures of slope of the psychometric function. *Percept Psychophys.* 2001;63:1348–1355.

21. Bach M. The Freiburg visual acuity test-variability unchanged by post-hoc re-analysis. *Graefes Arch Clin Exp Ophthalmol.* 2007;245:965–971.

22. Carkeet A, Bailey IL. Slope of psychometric functions and termination rule analysis for low contrast acuity charts. *Ophthalmic Physiol Opt.* 2017;37(2):118–127.

23. Hazel CA, Elliott DB. The dependency of LogMAR visual acuity measurements on chart design and scoring rule. *Optom Vis Sci.* 2002;79(12):788–792.

24. Lesmes LA, Dorr M. Active learning for visual acuity testing. In: *Proceedings of the 2nd International Conference on Applications of Intelligent Systems.* APPIS '19. New York, NY: Association for Computing Machinery; 2019:1–6.

25. Zhao Y, Lesmes LA, Dorr M, Bex PJ, Lu Z-L. Psychophysical validation of a novel active learning approach for measuring the visual acuity behavioral function. *Trans Vis Sci Tech.* 2021;10:1.

26. Lu Z-L, Dosher BA. *Visual Psychophysics: From Laboratory to Theory.* Cambridge, MA: MIT Press; 2013.

27. Clayton D, Hills M. *Statistical models in epidemiology.* Oxford, UK: Oxford University Press; 1993.

28. Edwards W, Lindman H, Savage LJ. Bayesian statistical inference for psychological research. *Psychol Rev.* 1963;70(3):193–242.

29. Wichmann FA, Hill NJ. The psychometric function: I. Fitting, sampling, and goodness of fit. *Percept Psychophys.* 2001;63:1293–1313.

30. Wichmann FA, Hill NJ. The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Percept Psychophys.* 2001;63:1314–1329.

31. Hou F, Lesmes L, Bex P, Dorr M, Lu Z-L. Using 10AFC to further improve the efficiency of the quick CSF method. *J Vis.* 2015;15:2.

32. Bailey IL, Lovie JE. New design principles for visual acuity letter charts. *Am J Optom Physiol Opt.* 1976;53:740–745.

33. Beck RW, Moke PS, Turpin AH, et al. A computerized method of visual acuity testing: Adaptation of the early treatment of diabetic retinopathy study testing protocol. *Am J Ophthalmol.* 2003;135:194–205.

34. Ferris FL, Kassoff A, Bresnick GH, Bailey I. New visual acuity charts for clinical research. *Am J Ophthalmol.* 1982;94:91–96.

35. Bach M. The Freiburg visual acuity test–automatic measurement of visual acuity. *Optom Vis Sci.* 1996;73:49–53.

36. Arditi A, Cagenello R. On the statistical reliability of letter-chart visual acuity measurements. *Invest Ophthalmol Vis Sci.* 1993;34:120–129.

37. Bokinni Y, Shah N, Maguire O, Laidlaw DAH. Performance of a computerised visual acuity measurement device in subjects with age-related macular degeneration: comparison with gold standard ETDRS chart measurements. *Eye.* 2015;29:1085–109.

38. Elliott DB, Sheridan M. The use of accurate visual-acuity measurements in clinical anti-cataract formulation trials. *Ophthalmic Physiol Opt.* 1988;8:397–401.

39. Laidlaw DAH, Abbott A, Rosser DA. Development of a clinically feasible logMAR alternative to the Snellen chart: Performance of the "'compact reduced logMAR'" visual acuity chart in amblyopic children. *Br J Ophthalmol.* 2003;87(10):1232–1234.

40. Laidlaw DAH, Tailor V, Shah N, Atamian S, Harcourt C. Validation of a computerised logMAR visual acuity measurement system (COMPlog): comparison with ETDRS and the electronic ETDRS testing algorithm in adults and amblyopic children. *Br J Ophthalmol.* 2008;92:241–244.

41. Patel PJ, Chen FK, Rubin GS, Tufail A. Intersession repeatability of visual acuity scores in age-related macular degeneration. *Invest Ophthalmol Vis Sci.* 2008;49:4347–4352.

42. Rosser DA, Laidlaw DAH, Murdoch IE. The development of a "reduced logMAR" visual acuity chart for use in routine clinical practice. *Br J Ophthalmol.* 2001;85:432–436.

43. Vanden Bosch ME, Wall M. Visual acuity scored by the letter-by-letter or probit methods has lower retest variability than the line assignment method. *Eye.* 1997;11:411–417.

44. Daniels M J, Kass RE. Nonconjugate Bayesian estimation of covariance matrices and its use in hierarchical models. *J Am Stat Assoc.* 1999;94(448):1254

45. Klotzke K, Fox J-P. Bayesian covariance structure modeling of responses and process data. *Front Psychol.* 2019;10:1675.

46. Thall PF, Wathen JK, Bekele BN, Champlin RE, Baker LH, Benjamin RS. Hierarchical Bayesian approaches to phase II trials in diseases with multiple subtypes. *Stat Med.* 2003;22(5):763–780.

47. Wang C, Lin X, Nelson KP. Bayesian hierarchical latent class models for estimating diagnostic accuracy. *Stat Methods Med Res.* 2020;29(4):1112–1128.

48. Yang J, Zhu H, Choi T, Cox DD. (2016). Smoothing and mean-covariance estimation of functional data with a Bayesian hierarchical model. *Bayesian Anal.* 2011;11(3):649–670.

49. Kruschke JK. *Doing Bayesian data analysis: a tutorial with R, JAGS, and Stan.* San Diego, CA: Academic Press; 2014.

50. Lee MD. A hierarchical Bayesian model of human decision-making on an optimal stopping problem. *Cogn Sci.* 2006;30(3):1–26.

51. Lee MD. How cognitive modeling can benefit from hierarchical Bayesian models. *J Math Psychol.* 2011;55(1):1–7.

52. Rouder JN, Lu J. An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychon Bull Rev.* 2005;12(4):573–604

53. Wilson JD, Cranmer S, Lu Z-L. A hierarchical latent space network model for population studies of functional connectivity. *Comput Brain Behav.* 2020;3:384–399.

54. Klotzke K, Fox J-P. Bayesian covariance structure modeling of responses and process data. *Front Psychol.* 2019;10:1675.

55. Lee MD. Three case studies in the Bayesian analysis of cognitive models. *Psychon Bull Rev.* 2008;15(1):1–15.

56. Thrane E, Talbot C. An introduction to Bayesian inference in gravitational-wave astronomy: Parameter estimation, model selection, and hierarchical models. *Publ Astron Soc Aust.* 2019;36:e010.

57. Reum JCP, Hovel RA, Greene C.M. Estimating continuous body size-based shifts in delta N-15-Delta C-13 space using multivariate hierarchical models. *Mar Biol.* 2015;162(2):469–478.

58. Wikle CK. Hierarchical Bayesian models for predicting the spread of ecological processes. *Ecology.* 2003;84(6):1382–1394.

59. Storz JF, Beaumont MA. Testing for genetic evidence of population expansion and contraction: An empirical analysis of microsatellite DNA variation using a hierarchical Bayesian model. *Evolution.* 2002;56(1):154–166.

60. Ahn W-Y, Krawitz A, Kim W, Busmeyer JR, Brown JW. A model-based FMRI analysis with hierarchical Bayesian parameter estimation. *J Neurosci Psychol Econ.* 2011;4(2):95–110.

61. Lee TS, Mumford D. Hierarchical Bayesian inference in the visual cortex. *J Opt Soc Am A Opt Image Sci Vis.* 2003;20(7):1434–1448.

62. Merkle EC, Smithson M, Verkuilen J. Hierarchical models of simple mechanisms underlying confidence in decision making. *J Math Psychol.* 2011;55(1):57–67.

63. Molloy MF, Bahg G, Li X, Steyvers M, Lu Z-L, Turner BM. Hierarchical Bayesian analyses for modeling BOLD time series data. *Comput Brain Behav.* 2018;1(2):184–213.

64. Molloy MF, Bahg G, Lu Z-L, Turner BM. Individual differences in the neural dynamics of response inhibition. *J Cogn Neurosci.* 2019;31(12):1976–1996.

65. Rouder JN, Sun DC, Speckman PL, Lu J, Zhou D. A hierarchical Bayesian statistical framework for response time distributions. *Psychometrika.* 2003;68(4):589–606.

66. Pérez GM, Archer SM, Artal P. Optical characterization of Bangerter foils. *Invest Ophthalmol Vis Sci.* 2010;51:609–613.

67. Odell NV, Leske DA, Hatt SR, Adams WE, Holmes JM. The effect of Bangerter filters on optotype acuity, Vernier acuity, and contrast sensitivity. *J AAPOS.* 2008;12(6):555–559.

68. Plummer M. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing.* 2003. Available at: https://www.r-project.org/nosvn/conferences/DSC-2003/Drafts/Plummer.pdf.

69. R Core Team. *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing; 2003. Available at: https://www.R-project.org/.

70. Gelman A., Rubin D. B. Inference from iterative simulation using multiple sequences, *Stat Sci.* 1992;7:457–511.

71. Ando T. Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models. *Biometrika.* 2007;94:443–458.

72. Ando T. Predictive Bayesian model selection. *Am J Math Manag Sci.* 2011;31:13–38.

73. Csaky K, Ferris F, Chew EY, Nair P, Cheetham JK, Duncan JL. Report from the NEI/FDA

endpoints workshop on age-related macular degeneration and inherited retinal diseases. *Invest Ophthalmol Vis Sci.* 2017;58(9):3456–3463.

74. Green DM, Swets JA. *Signal Detection Theory and Psychophysics*. New York, NY: John Wiley & Sons; 1966.

75. Ashby FG, Townsend JT. Varieties of perceptual independence. *Psychol Rev.* 1986;93(2):154–179.

76. Borm GF, Fransen J, Lemmens WAJG. A simple sample size formula for analysis of covariance in randomized clinical trials. *J Clin Epidemiol.* 2007;60(12):1234–1238.

77. Egbewale BE, Lewis M, Bias Sim J., Precision and statistical power of analysis of covariance in the analysis of randomized trials with baseline imbalance: A simulation study. *BMC Med Res Methodol.* 2014;14:49.

78. Wilcox R. *Modern Statistics for the Social and Behavioral Sciences: A Practical Introduction* (pp. 101–102). Boca Raton, FL: CRC Press; 2012.

79. Gu H, Kim W, Hou F, et al. A hierarchical Bayesian approach to adaptive vision testing: A case study with the contrast sensitivity function. *J Vis.* 2016;16(6):15.

80. Kim W, Pitt MA, Lu Z-L, Steyvers M, Myung JI. A hierarchical adaptive approach to optimal experimental design. *Neural Comput.* 2014;26(11):2465–2492.