Accounting for Standard Errors of Vision-Specific Latent Trait in Regression Models

Wan Ling Wong1,2 Xiang Li1,3,4 Jialiang Li3 Tien Yin Wong1,2,5 Ching-Yu Cheng1,2,4,5 and Ecosse L. Lamoureux1,5

1Singapore Eye Research Institute, Singapore National Eye Centre, Singapore
2Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore
3Department of Statistics and Applied Probability, National University of Singapore, Singapore
4Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, New York, United States
5Office of Clinical Sciences, Duke-NUS Graduate Medical School, Singapore
6Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore

Correspondence: Ecosse L. Lamoureux, Singapore Eye Research Institute, Singapore National Eye Centre, 11 Third Hospital Lane, Singapore 168751; ecosse.lamoureux@seri.com.sg.
Submitted: February 19, 2014 Accepted: June 10, 2014

PURPOSE. To demonstrate the effectiveness of Hierarchical Bayesian (HB) approach in a modeling framework for association effects that accounts for SEs of vision-specific latent traits assessed using Rasch analysis.

METHODS. A systematic literature review was conducted in four major ophthalmic journals to evaluate Rasch analysis performed on vision-specific instruments. The HB approach was used to synthesize the Rasch model and multiple linear regression model for the assessment of the association effects related to vision-specific latent traits. The effectiveness of this novel HB one-stage “joint-analysis” approach allows all model parameters to be estimated simultaneously and was compared with the frequently used two-stage “separate-analysis” approach in our simulation study (Rasch analysis followed by traditional statistical analyses without adjustment for SE of latent trait).

RESULTS. Sixty-six reviewed articles performed evaluation and validation of vision-specific instruments using Rasch analysis, and 86.4% (n = 57) performed further statistical analyses on the Rasch-scaled data using traditional statistical methods; none took into consideration SEs of the estimated Rasch-scaled scores. The two models on real data differed for effect size estimations and the identification of “independent risk factors.” Simulation results showed that our proposed HB one-stage “joint-analysis” approach produces greater accuracy (average of 5-fold decrease in bias) with comparable power and precision in estimation of associations when compared with the frequently used two-stage “separate-analysis” procedure despite accounting for greater uncertainty due to the latent trait.

CONCLUSIONS. Patient-reported data, using Rasch analysis techniques, do not take into account the SE of latent trait in association analyses. The HB one-stage “joint-analysis” is a better approach, producing accurate effect size estimations and information about the independent association of exposure variables with vision-specific latent traits.

Keywords: measurement error, latent trait, Rasch analysis
and regression models using the Hierarchical Bayesian (HB) approach that accounts for SEs of latent trait. This modeling is appropriate for the assessment of association effects related to the vision-specific latent trait, with proper treatment of its associated SEs for a more accurate and contemporary estimates of association effects. We compared the one-stage “joint-analysis” and two-stage “separate-analysis” model results using real data and assessed the performance of the methods in a simulation study based on the frequently used Andrich rating scale model.26

**Materials and Methods**

**Literature Review**

We systematically reviewed publications that used Rasch analysis by searching the electronic databases of PubMed in the top four Ophthalmic journals (Ophthalmology, American Journal of Ophthalmology [AJO], British Journal of Ophthalmology [BJO], and IOVS) for relevant papers published up to July 2013, with the following search terms (formatted for PubMed search):


The strategy identified 70 articles and the full texts were reviewed (by WLW and XL) to identify studies having performed Rasch analysis on visual functioning questionnaire data. Of the 70 articles identified, the following were excluded: two letters,27,28 one study that applied the Rasch model to investigate inter-reader agreement,29 and another study that focused on the genetic components of the optic nerve head.30

The remaining 66 articles related to visual functioning data were reviewed for choice of Rasch model, implementation software, and sample size of studies (Table 1).

**Pitfalls in Observed Analysis Framework (Two-Stage “Separate-Analysis” Procedure)**

All 66 articles reviewed performed Rasch analysis to evaluate the validity, reliability, and measurement characteristics of instruments (i.e., visual functioning) for their population sample data. Most (86.4%) performed further statistical analysis on the Rasch-scaled score (i.e., vision-specific latent trait), such as performing correlations or linear regressions with visual acuity, demographic, or clinical data to assess the impact of visual impairment and other patients’ characteristics or factors on visual functioning.

However, none of the articles mentioned or discussed the potential bias in estimation of association effects and the understimation of their SEs31 having ignored the associated uncertainties involved in the estimation of the latent trait when used naively in subsequent association analysis. In the first stage, the vision-specific latent trait (i.e., vision functioning) was modeled and estimated given a set of item responses using a polytomous Rasch model (e.g., Andrich rating scale model when item response options are more than dichotomous, that is, three or more). In the second stage, relationships between the estimated Rasch-scaled data (treated as known outcome variable) and risk factors were analyzed using regression techniques. Ignoring the uncertainty regarding the abilities within the regression model may lead to biased estimation of association effects.31 Underestimation of SEs may also result in false identified positive factors and hence mislead statistical inferences.31

Moreover, a key assumption in the Rasch model states that a change in the latent variable is completely described by the item characteristic functions (the relationship of the latent trait and responses of the items) and hence any association analysis on the latent trait with other covariates performed in the second stage can violate and contradict the key assumptions in the first stage of Rasch analysis (having assumed that vision-specific latent trait only depends on item response data). Such
estimation procedure can cause serious underestimation of the SEs of the model parameters.\textsuperscript{34}

Our literature review also observed that studies assessing visual-functioning traits were often conducted for moderately small sample sizes (median 240 with interquartile range of 497) and, together with response data (and Rasch-scaled scores) that are typically non-normally distributed, it is precarious to rely on asymptotic approximations and the properties of conditional maximum likelihood estimates obtained from analysis software without showing them to be accurate.\textsuperscript{52} Similarly, validation inference based on correlations may not be accurate.

Our study analysis and discussion focuses on the Andrich rating scale\textsuperscript{26} because it is the most frequently used polychotomous Rasch model for vision-specific instruments.

### Andrich Rating Scale Model

The Andrich rating scale\textsuperscript{26} model is an extension of the Rasch model\textsuperscript{53} for polychotomous responses (i.e., three or more categories). The natural log of the likelihood ratio of adjacent response category probabilities is given by

\[
\ln(P_{ik,y}/P_{ik,(y-1)}) = \theta_i - \eta_k - \gamma_y,
\]

where \(P_{ik,y}\) is the probability of person \(i\) on encountering item \(k\) would be observed in category \(y\), \(\theta_i\) is ability trait for the \(i\)th person (i.e., \(i = 1, \ldots, n\)), \(\eta_k\) is the item \(k\) difficulty parameter (i.e., \(k\) items means \(k = 1, \ldots, k\)), and \(\gamma_y\) is the threshold for category \(y\) (e.g., items with five categories means \(y\) is in the range of integers 1 to 5), which is constant across items.

### Proposed Analysis Framework (One-Stage “Joint-Analysis”)

A rigorous alternative is to combine the observed two-stage analysis procedure to overcome the problematic issues described above. Item response data structure are hierarchical, as item responses are nested within respondents and respondents also may be nested (e.g., patients nested in hospitals). Such relationships can be adequately explained by the multilevel Rasch model using the HB approach. Model parameters also can be incorporated and estimated from the item response data without having to condition on estimated person ability parameter (i.e., latent trait). In Supplementary Figure S1, a path diagram of this multilevel Rasch model is depicted and explained.

For example, the combined model for linear regression (i.e., continuous latent trait outcome) can be written as

\[
\begin{align*}
\ln(P_{ik,y}/P_{ik,(y-1)}) & = \theta_i - \eta_k - \gamma_y, \\
\theta_i & = \beta \times X_i + \epsilon_i,
\end{align*}
\]

where \(\beta\) is the beta coefficient from linear regression, \(X_i\) is the observed covariates of person \(i\), and \(\epsilon_i\) is the residual random error. The HB approach provides an elegant execution of the multilevel Rasch modeling framework that allows the incorporation of explanatory variables or covariates at different levels of hierarchy by specifying parameters to come from a specific distribution with parameters and possibly hyper-parameters that are, themselves, estimated prior information.\textsuperscript{53} All model parameters can then be estimated simultaneously using the Monte Carlo Markov Chain method with the JAGS program.\textsuperscript{54,56} The proposed procedure enables direct estimation of beta coefficients for association effects without having to explicitly know the latent trait measurements (i.e., personal ability). The JAGS codes used to fit our example model in one-stage “joint-analysis” HB approach are provided in the Supplementary Material, and can be altered readily to conform to different data structure.

### Comparison of Methods Using Real Data

Both the HB one-stage “joint-analysis” and the two-stage “separate-analysis” methods were performed to assess the relationship of reading and writing literacy on visual functioning (measured by a modified VF-9 questionnaire) using data from the Singapore Malay Eye Study (SiMES),\textsuperscript{57} a population-based cross-sectional study of 3280 Singaporean Malays older than 40. Previous studies suggest an association of inadequate literacy with systemic health and hence the influence of literacy on vision functioning (in addition to visual impairment), another aspect contributing to vision-specific quality of life, is important.\textsuperscript{58,59} Association of reading and writing literacy with visual functioning were adjusted for age, sex, language of interview, body mass index (BMI), occupation, marital status, income, housing type, education, smoking status, and presenting visual acuity in the better-seeing eye.

### Simulation Study

As there is no “gold standard” in the comparison of methods using real data, we conducted a simulation study to demonstrate the performance of our proposed HB one-stage “joint-analysis” approach as compared with the observed two-stage “separate-analysis” procedure. Two independent covariates (\(X_{1}, X_{2}\)), a continuous variable data, such as standardized age, and a binary variable, such as sex, were simulated with prespecified association effects (\(\beta_1, \beta_2\)) for the impact of these two covariates with the latent visual functioning ability parameter and, hence, these were considered as the “true” association effects or “gold standard” for reference when we re-run analysis on our simulated data using both analytical methods. Association estimates and their SEs from both approaches were computed to assess their performance, where estimates closer to the “gold standard” indicate higher accuracy and smaller SEs suggest greater precision. The calibration of nine item difficulty parameters, \(\eta_k\) was fixed according to table 3 of a study conducted by Lamoureux et al.\textsuperscript{49} that performed a systematic evaluation of the reliability and validity of the visual functioning questionnaire (VF-11) using Rasch analysis that was later modified to nine items (VF-9) to tailor fit to the Asian population. We also investigated the empirical power for both approaches. We provide a detailed description of our simulation study in the Supplementary Material.

### Results

Table 1 shows the summary of articles reviewed in the four major ophthalmic journals that have performed Rasch analysis for visual functioning-related instrument data. Most (65.1%) performed Rasch analysis by using the Andrich rating scale model and most (68.2%) conducted Rasch analysis by using Winsteps software (Winsteps, Beaverton, OR, USA). The median sample size of these studies was 240 with interquartile range of 497.

Associations of inadequate reading and writing literacy with visual functioning adjusted for potential confounding variables (models 1 and 2, respectively) derived from both approaches analyzed on our simulated data are shown in Table 2. Comparison of both approaches for Model 1 assessing reading literacy showed no difference in statistically significant factors identified but results were not consistent for writing literacy in Model 2. Inadequate writing was statistically significant based
### Table 2. Comparison Between Approaches Using Real Data*

<table>
<thead>
<tr>
<th>Approach</th>
<th>Model 1 (Reading Literacy)</th>
<th>Model 2 (Writing Literacy)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2-Stage</td>
<td>P</td>
</tr>
<tr>
<td>Age, y</td>
<td>0.000 (−0.008 to 0.008)</td>
<td>0.906</td>
</tr>
<tr>
<td>Sex</td>
<td>0.012 (−0.161 to 0.184)</td>
<td>0.894</td>
</tr>
<tr>
<td>Language of interview</td>
<td></td>
<td></td>
</tr>
<tr>
<td>English vs. Malay</td>
<td>−0.227 (−0.501 to 0.046)</td>
<td>0.103</td>
</tr>
<tr>
<td>Others vs. Malay</td>
<td>0.569 (−0.685 to 2.002)</td>
<td>0.336</td>
</tr>
<tr>
<td>BMI</td>
<td>0.002 (−0.010 to 0.014)</td>
<td>0.759</td>
</tr>
<tr>
<td>Occupation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Office work</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Service work</td>
<td>−0.060 (−0.263 to 0.143)</td>
<td>0.559</td>
</tr>
<tr>
<td>Factory work</td>
<td>−0.186 (−0.451 to 0.078)</td>
<td>0.166</td>
</tr>
<tr>
<td>Homemaking</td>
<td>−0.118 (−0.376 to 0.141)</td>
<td>0.571</td>
</tr>
<tr>
<td>Unemployed/other</td>
<td>−0.091 (−0.334 to 0.152)</td>
<td>0.463</td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never married</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Married</td>
<td>0.072 (−0.165 to 0.309)</td>
<td>0.552</td>
</tr>
<tr>
<td>Separated/divorced</td>
<td>−0.218 (−0.527 to 0.091)</td>
<td>0.167</td>
</tr>
<tr>
<td>Widowed</td>
<td>0.082 (−0.237 to 0.400)</td>
<td>0.615</td>
</tr>
<tr>
<td>Income</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; SGD$1000/mo</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>&lt; SGD$1000/mo</td>
<td>0.049 (−0.114 to 0.212)</td>
<td>0.556</td>
</tr>
<tr>
<td>Retirement income</td>
<td>0.065 (−0.127 to 0.256)</td>
<td>0.508</td>
</tr>
<tr>
<td>Current housing status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/2-room flat</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>3/4-room flat</td>
<td>−0.259 (−0.416 to −0.101)</td>
<td>0.001</td>
</tr>
<tr>
<td>5-room/private house</td>
<td>−0.281 (−0.477 to −0.085)</td>
<td>0.005</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No formal education</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Primary education</td>
<td>−0.164 (−0.394 to 0.066)</td>
<td>0.162</td>
</tr>
<tr>
<td>Secondary education</td>
<td>−0.152 (−0.411 to 0.107)</td>
<td>0.249</td>
</tr>
<tr>
<td>Poly/University</td>
<td>−0.124 (−0.427 to 0.180)</td>
<td>0.424</td>
</tr>
<tr>
<td>Smoking status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Past or never</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Current</td>
<td>−0.203 (−0.359 to −0.048)</td>
<td>0.010</td>
</tr>
<tr>
<td>PVA of better eye</td>
<td>−0.934 (−1.165 to −0.703)</td>
<td>0.000</td>
</tr>
<tr>
<td>Read</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes vs. no</td>
<td>0.297 (0.026 to 0.568)</td>
<td>0.052</td>
</tr>
<tr>
<td>Write</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes vs. no</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Based on SiMES data.*
on the HB one-stage “joint-analysis” but was not significant in the two-stage “separate-analysis” approach. Smaller association effects also were estimated in the two-stage “separate-analysis” approach, which is consistent with the attenuation bias observed in the simulation study.

Simulation results (based on sample size of 300) on the association effects and their SEs of a continuous measurement, such as standardized age ($\beta_1$) and that for a binary factor, such as sex ($\beta_2$), with the vision-specific latent trait compared with the “gold standards” for both approaches are depicted in the Figure. There is greater inaccuracy (average of 5-fold increase in bias) in effect size estimations from the frequently used two-stage “separate-analysis” procedure compared with the proposed HB one-stage “joint-analysis” approach. We also observed an attenuation bias in estimations (shrunk toward zero) from the two-stage procedure. Smaller SEs for estimates were expected for the two-stage procedure having assumed no uncertainty in the latent trait measurements, but the slightly larger SEs (average of one-tenth-fold difference) from the HB approach suggest comparable precision for a more accurate estimation of associations. The impact of sample size on the level of bias tabulated in Supplementary Table S1 shows marked decrease in bias for estimations from HB one-stage “joint-analysis” with increasing sample size (from 0.011 to 0.001 and $-0.010$ to $-0.003$ for betas 0.2 and 0.5, respectively, when sample size increases from 50 to 600). Bias is negligible with sufficiently large sample size, such as 100 or more, in our study. However, the two-stage “separate-analysis” approach consistently produced negative bias regardless of sample size (little difference in bias with sample sizes from 50 to 600). Hence, bias cannot be eliminated/reduced even with large sample size. The negative bias reflects the attenuation bias observed in the Figure. Furthermore, the level of power in our proposed HB approach was comparable to the two-stage approach (no difference for beta $\geq 0.5$ at 5% significance level and less than one-twelfth-fold difference for beta at 0.2 for both continuous and categorical variables), despite taking into account uncertainty in the latent trait (Supplementary Table S2).

**DISCUSSION**

This is the first study to assess the performance of two regression models for visual functioning, comparing the frequently used two-stage “separate-analysis” method (ignoring SE of the dependent latent trait) and our proposed one-stage “joint-analysis” approach in terms of estimation accuracy of association effects, precision of their SEs, and power. Association effect sizes from our real data analysis were observed to be smaller in the two-stage “separate-analysis” approach with slightly tighter intervals and the identification of significant factors between approaches were different. Our simulation study results (assessing methods performance) provided support for these observations. Attenuation bias, the shrinking of estimations toward zero, was found using the two-stage procedure (a phenomenon expected from ordinary least-squares regression of explanatory variables with measurement errors) that explains the (artificial) smaller association effects observed in our real-data analysis. This attenuation bias exists regardless of sample size.

The one-stage “joint-analysis” approach allows the estimation of all model parameters simultaneously and hence integrates visual functioning in the regression model accounting for its SEs. Simulation results also showed that the one-stage “joint-analysis” method produced highly accurate estimations (average of 5-fold decrease in bias), with comparable precisions and power as compared to the commonly used two-stage “separate-analysis” procedure. The magnitude of SE also affects the size of attenuation bias. Accurate estimation of effect size and its variance are both critical to statistical significance testing results that directly influence our interpretation of risk factors. Hence, moving forward, the one-stage “joint-analysis” approach is preferred when we perform regression analysis with dependent outcome that is essentially a latent variable being derived from some prior analyses. Such analytical frameworks were already adopted to assess impact on longitudinal vision outcome in terms of vision-related quality of life using multilevel item response models,41,42 and similarly for some other diseases.43,44

Our model codes provided in the Supplementary Material can be altered readily to analyze data from any instruments/questionnaires with any choice of Rasch or item response theory models (e.g., Partial Credit model, Graded Response model) and complexity of multilevel regression models as depicted in Supplementary Figure S1 to adequately describe the hierarchical data structure, incorporate different sources of uncertainty, and inclusion of explanatory covariates at different levels. Furthermore, latent variables also can be analyzed as independent covariates (as required) instead of an outcome variable used in our simulation example.

In our literature review, visual functioning rating scale instruments/questionnaires were mainly validated using Rasch analysis for their well-known scaling and measurement properties. Without a realist interpretation of latent variables, actual Rasch-scaled scores do not have straightforward meaning and its interpretations are based on relative comparisons of the scaled scores (i.e., relative difference tells us how much more of persons’ visual functioning ability compared with another). Many reviewed articles provide ready-to-use spreadsheets that convert raw scores entered to Rasch-scaled scores for their respective instruments to benefit clinicians and researchers unfamiliar with Rasch analysis who may wish to use its scoring benefits. It is important to note that the population of respondents plays an important part in the
probability model for each response and so the personal ability and item parameters always will be estimated with respect to a population. Hence, ready-to-use spreadsheets should be used only on different samples of individuals from the same population as validated in the article and that Rasch-scaled scores are not comparable between studies from different populations unless it happens (rarely) that both populations have identical item characteristic functions. Researchers also were unaware of the SEs associated with Rasch-scaled scores and performed further analysis directly with simple statistical tests, such as independent t-tests to examine between-group differences in the instrument scores across various sociodemographic variables and levels of vision impairment.

The strength of our study includes the simulating datasets with various prespecified association effects that act as “ground truth” to enable the comparison between methods. The simulating conditions in our study were, however, limited to only the Andrich Rating Scale model with linear regression analysis for covariates at the respondents’ level. Statistical computations are necessary for applying our proposed analysis using the HB approach and some background in statistics and programming skills are needed to alter codes to conform to other conditions.

In conclusion, there is a need to account for SEs associated with vision-specific latent trait in association analysis. We demonstrated that our HB one-stage “joint-analysis” approach is a better method that produces greater accuracy with comparable power and precision in estimation of association effects compared with the frequently used two-stage procedure, despite taking into account greater uncertainty due to the latent trait. The study finding has direct implications in our inference drawn from statistical significance of risk factors.

Acknowledgments
Supported by Grant 0796/2003 from the National Medical Research Council, Singapore.

Disclosure: W.L. Wong, None; X. Li, None; J. Li, None; T.Y. Wong, None; C.-Y. Cheng, None; E.L. Lamoureux, None

References


