Appl. Statist. (2002)
51, *Part* 1, *pp.* 37–57

Building marginal models for multiple ordinal measurements

Guan-Hua Huang,

University of Wisconsin, Madison, USA

Karen Bandeen-Roche

Johns Hopkins University, Baltimore, USA

and Gary S. Rubin

University College London, UK

[Received May 1999. Final revision June 2001]

Summary. Biomedical and psychosocial researchers increasingly utilize multiple indicators to assess an outcome of interest. We apply the ordinal estimating equations model for analysing this kind of measurement. We detail the special complexities of using this model to analyse clustered non-identical items and propose a workable model building strategy. Three graphical methods—cumulative log-odds, partial residual and Pearson residual plotting—are developed to diagnose the adequacy of models. The benefit of incorporating interitem associations and the trade-off between simple *versus* complex models are evaluated. Throughout the paper, an analysis to determine how measured impairments affect visual disability is used for illustration.

Keywords: Generalized estimating equation; Global odds ratio; Graphical diagnosis; Model building; Partial residual plot; Proportional odds model

1. Introduction

Increasingly in biomedical studies, health status is inferred through multiple indicators. For example, physical disability in older people is often quantified as categorized responses to a series of questions about their ability to perform routine tasks of living (Katz *et al*., 1963; Lawton and Brody, 1969). Similarly, visual disability may be quantified as responses about ability to perform various visual tasks (Steinberg *et al*., 1994). This paper reports our work in tailoring generalized estimating equation (GEE) methodology for analysing such 'multiply measured' outcomes, as motivated by the need to analyse visual functioning data recently collected as part of the Salisbury eye evaluation (SEE) project (West *et al*., 1997).

Broadly, we aim to describe the dependence of multiply measured outcomes on explanatory variables. Researchers commonly attempt this by adding outcomes indicators into scores and then regressing the scores on covariates (e.g. Stewart and Ware (1992)), or they use latent variable models to infer unobservable scores that underlie the observed responses (e.g. Bartholomew (1987) and Bandeen-Roche *et al*. (1997)). However, the scoring method risks combining indicators of distinct processes and hence masks associations

Address for correspondence: Guan-Hua Huang, Department of Population Health Sciences, Medical School, University of Wisconsin, 610 Walnut Street, Madison, WI 53705, USA.
E-mail: guanhuahuang@facstaff.wisc.edu

 0035–9254/02/51037

between outcomes and risk factors, and latent variable models come at the price of strong modelling assumptions which may critically influence analytic findings. In contrast, an investigator might individually analyse how the separate indicators are related to variables of interest. However, this approach wastes information for estimating parameters describing mean relationships between outcomes and predictors, because a person's multiple responses may be subject to shared influences that are unexplained by measured explanatory variables (e.g. person-specific measurement conditions and individual robustness or frailty). Such influences induce correlations between individuals' indicator-specific deviations from mean predicted outcomes. Analysing measurements separately ignores information for extracting deviations that are shared across indicators from those that distinguish one person from another, resulting in inefficient estimators of predictor relationships with responses. Precision gains from utilizing information across indicators may in fact be substantial (Liang and Zeger, 1993), which may be crucial in applications where data are costly (laboratory science), difficult (rare diseases) or burdensome (gerontology) to obtain.

One well-discussed alternative to analysing the components of multiply measured outcomes separately is to model the component response means individually, but to incorporate the association between multiple responses into the estimation procedure to achieve precise inferences. Methods that implement this strategy range from classical multivariate regression and analysis of variance (e.g. Morrison (1990)), weighted least squares (Jacquez *et al.*, 1968), seemingly unrelated regressions (Zellner, 1962) and marginal models (Liang and Zeger, 1986; Zhao and Prentice, 1990; Fitzmaurice and Laird, 1993) to random-effects models (Scheffé, 1959; Laird and Ware, 1982; Bryk and Raudenbush, 1992). Among these, marginal models have achieved particular popularity in quantitative biomedical settings. To date, though, literature and available software to implement these methods have been primarily concerned with within-person associations due to repeated measurements of a single response, which assumes equal associations between outcomes and covariates across different indicators. In contrast, the multiply measured response setting tends to require more complex models—because covariates may affect various indicators differently, and because simple models for the pattern of associations between indicators may be inadequate.

This paper proposes procedures for fitting marginal models to ordinal multiple-response functioning data, illustrating with a detailed example throughout. Our work has three implications. First, we tailor parameterizations of marginal models to describe multiply measured responses, while retaining the benefits of precision of estimating responses jointly. Second, the method proposed allows constraining covariate effects to be equal across indicators, which achieves the summary sought by scoring and latent variable models and provides a framework for testing whether such a summary reasonably captures the complexity of the data. Third, we propose graphical displays to assess and refine model adequacy. Cumulative log-odds and Pearson residual plots, which modify the diagnostics that are commonly used in ordinary linear regression, are used to evaluate the proportional odds assumption and overall fit. A new partial residual plot for multivariate ordinal responses is formulated to check the functional relationship between response variables and covariates.

The remainder of the paper is organized as follows. In Section 2, we briefly describe the research project that motivated this study, overview the statistical method that we shall apply (the Heagerty–Zeger (HZ) model; Heagerty and Zeger (1996)) and propose models for multiply measured outcomes. Section 3 proposes a strategy to build HZ models for multiple-indicator data. This strategy is designed to allow different covariate effects on different outcome components while managing the computational intensity of applying the HZ model. It also describes our graphical displays for checking the model fit. In Section 4, we explore the

potential gain in precision from accounting for association in estimating regression coefficients and consider the trade-off between complex and simple mean models. Section 5 concludes by discussing the possible generalization of the model proposed and the potential for impossible predicted responses.

2. Background

2.1. Research application

The SEE project is a population-based prospective study of how vision affects functioning in older people. The study has been described in detail elsewhere (West *et al.*, 1997). In brief, potential study participants were selected as an age-stratified random sample of people who were 65–84 years old living in the community in certain Salisbury, Maryland, zip codes at a given sampling date. Sampled individuals who could communicate in English, travel to a central clinic site and score higher than 17 points on the mini-mental state examination (Folstein *et al.*, 1975) were designated as eligible to participate in the study. 2520 people agreed to participate in the home and clinic components of the study.

The analysis that we report here aims to describe the associations between self-reported ability to do vision-related tasks and distinct measured aspects of vision, adjusting for potentially confounding variables. Such a description has significant implications for designing interventions to minimize task disability among the visually impaired. In the SEE project, each participant's ability to do specific visually oriented activities was determined by using the activities of daily vision scale (ADVS) questionnaire (Mangione *et al.*, 1992). For each activity, participants reported five levels of difficulty in doing the activity, which we condensed to four: 1, unable to do or extreme difficulty because of poor vision; 2, moderate difficulty; 3, a little difficulty; 4, no difficulty. Prior work arranges the activities into five subscales that measure self-reported difficulty in doing tasks related to *near* and *far* vision, *day* and *night* driving, and situations with *glare*. In this paper, we shall focus on the far vision subscale, which assesses five activities: reading street signs in daylight (daysgn), reading street signs at night (nightsgn), walking down steps during daylight (daystp), walking down steps in dim light (dimstp) and watching television (tvwatch).

The SEE measures of vision have been described in detail elsewhere (Rubin *et al.*, 1997). In brief, these included

(a) binocular visual acuity at regular luminance (vabnor), which is scored as the total number of letters read correctly in a standard eye chart and then converted to log(MAR) (minimum resolvable angle; Bailey *et al.* (1991)),
(b) contrast sensitivity of the better eye (bcs), which is scored as the number of letters read in an eye chart whose symbols grow fainter rather than smaller (Pelli *et al.*, 1988),
(c) sensitivity to glare of the worse eye (difcs), whose score is the number of letters on the contrast sensitivity chart correctly identified without glare minus the number of letters identified with glare added,
(d) binocular stereoacuity (logster), which measures an important aspect of depth perception as a threshold in log-seconds of arc, using the Randot circles chart, and
(e) central binocular visual field (bestcen), which is scored as the number of missed points of light presented one at a time in the central 30° of the field.

For all the measures except contrast sensitivity, a higher score indicates worse vision. The null hypothesis as reflected in current clinical practice is that impairments (b)–(e) are unassociated with self-reported task disability after adjusting for visual acuity.

The potential confounding variables selected for analysis were age at clinic examination (age), mini-mental state examination score (mmse), number of years of education (edu), an indicator of being female (sex), an indicator of being African-American (race), the number of reported comorbid diseases (comorbid) and the general health questionnaire depression subscale score (ghqdscor) (Goldberg, 1972). These have all been reported or hypothesized to affect self-reported functioning (Bandeen-Roche *et al.*, 1999).

2.2. Marginal regression models for clustered ordinal measurements

This paper presents a regression model with self-reported visual ability (ADVS items) as the outcome, and the measured visual impairments and potential confounding variables as covariates. Because the ADVS items are ordinal scales, the proportional odds model (McCullagh, 1980) is a good candidate to describe the mean relationship between item responses and predictors. We might fit five separate ordinal logistic regressions of the ADVS far vision items. However, this approach neither yields efficient parameter estimates nor provides a reasonable framework for comparing predictor associations between items. Such comparisons are important in our ophthalmologic setting to gain insight into specifically how the various vision measures relate to functioning. Thus, we estimated the item models jointly, accounting for item response associations within individuals by using ordinal estimating equations as proposed by Heagerty and Zeger (1996).

2.2.1. Model

Similarly to Lipsitz *et al.* (1991), Heagerty and Zeger (1996) proposed two distinct regressions for clustered ordinal responses—proportional odds models (McCullagh, 1980) to describe the marginal response means and global odds ratio models (Dale, 1986) to describe the marginal response–pair associations. Formally, let $\mathbf{O}_i = (O_{i1}, O_{i2}, \ldots, O_{in})$ represent a cluster of ordinal responses for the ith sampling unit, $i = 1, 2, \ldots, N$. In our application, $\mathbf{O}_i = (O_{i1}, O_{i2}, O_{i3}, O_{i4}, O_{i5})$ represents the collection of far vision item responses for the ith SEE participant, so O_{ij} is the level of difficulty reported by the ith participant doing the jth far vision activity. We assume that O_{ij} is realized at some $c \in [1, 2, \ldots, C]$. Heagerty and Zeger's framework represents this through cumulative indicator variables $Y_{ijc} = I(O_{ij} > c) = 1$ if $O_{ij} > c$ and $Y_{ijc} = 0$ otherwise, where $c \in [1, 2, \ldots, C - 1]$. $E(Y_{ijc}) = \Pr(Y_{ijc} = 1)$ defines the probability, say, that the ith participant reports better visual ability than level c in the jth far vision item.

The HZ model for the marginal means is

$$\text{logit}\{E(Y_{ijc})\} = \log\left\{\frac{\Pr(Y_{ijc} = 1)}{1 - \Pr(Y_{ijc} = 1)}\right\} = \theta_c + \mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta}, \tag{2.1}$$

where $\mathbf{x}_{ij}$ is the vector of covariates associated with O_{ij}, $i = 1, 2, \ldots, N$, $j = 1, 2, \ldots, n$, $c = 1, 2, \ldots, C - 1$. Specifically, the association between the outcome variable and covariates is assumed to be the same for each item ($\boldsymbol{\beta}_j = \boldsymbol{\beta}$, each j). In longitudinal or clustered measurements on a single item, model (2.1) is reasonable. In contrast, multiply measured outcomes consist of clustered but non-identical items. Covariates may affect individual measurements differently, and a more complex model is needed. For example, night driving is more visually demanding than driving during the day. At least, then, a reasonable model must accommodate different difficulty distributions for different items. Our model does this by using dummy variables indicating the item measured by a given response:

$$\text{logit}\{E(Y_{ijc})\} = \theta_c + \sum_{k=2}^{5} \gamma_k I_{kj},$$

where $i = 1, \ldots, 2520$, $j = 1, \ldots, 5$ (1, daysgn; 2, nightsgn; 3, daystp; 4, dimstp; 5, tvwatch), $c = 1, 2, 3$ and $I_{kj} = 1$ if $j = k$, or $I_{kj} = 0$ if $j \neq k$. The item effects γ_k are *log-odds ratios* for reporting greater difficulty on the kth *versus* the first far vision item—e.g. *factors* by which the odds of reporting difficulty greater than level c on items $k = 2, \ldots, 5$ exceed or are less than the odds of reporting that level of difficulty on item 1.

Turning to our treatment of explanatory variables, suppose that the goal is to determine the association between difficulty with far vision and gender. Equation (2.1) suggests the model

$$\text{logit}\{E(Y_{ijc})\} = \theta_c + \sum_{k=2}^{5} \gamma_k I_{kj} + \beta \text{ sex}_i, \tag{2.2}$$

where sex_i is 1 for females and 0 for males, which assumes that the odds ratio relating a far vision difficulty greater than level c to being female *versus* male ($\exp(\beta)$) is independent of j and c. This is not necessarily reasonable: for instance, women and men may report difficulty with walking down steps more differentially than with watching television. We allow such distinctions by adding item-by-sex interactions ($I_{kj} \times \text{sex}_i$ terms) to the design matrix:

$$\text{logit}\{E(Y_{ijc})\} = \theta_c + \sum_{k=2}^{5} \gamma_k I_{kj} + \beta \text{ sex}_i + \sum_{k=2}^{5} \tau_k (I_{kj} \times \text{sex}_i). \tag{2.3}$$

This formula also allows formal testing for whether $(\tau_2, \ldots, \tau_5) = \mathbf{0}$; a failure to reject this hypothesis would suggest that gender affects difficulty in reporting comparably across tasks. In this case, β would summarize the gender effect much as the gender coefficient in a regression of far vision subscale scores on covariates.

Finally, the proportional odds assumption in model (2.3) pools information by assuming a single regression function for all the derived binary responses Y_{ijc}, $c = 1, 2, \ldots, C-1$. However, in our example, the common wisdom is that women tend to report having moderate or a little difficulty ($c = 2, 3$) more readily than men do, but that they may not report great difficulty ($c = 1$) more readily than men do. Gender effects might be different across cutpoints c. We can correct this by adding cut-off by sex interactions to the design matrix. Similarly, cut-off by item interactions can be added to allow more flexible differences in item difficulty distributions than the *shifts* that are accommodated by the item main effects. The three-way interaction, cut-off by item by sex, is useful in specifying different gender effects across items and cutpoints. With all these interactions, model (2.3) may grow to include many terms. In fact, a model that is saturated in item, gender and cut-off interactions is equivalent to

$$\text{logit}\{E(Y_{ijc})\} = \theta_{jc} + \beta_{jc} \text{ sex}_i. \tag{2.4}$$

In contrast with model (2.3), this more general parameterization has the advantage of correctly accommodating outcomes whose items have different numbers of categories.

To describe the associations between responses from the same cluster, we use the global odds ratio (Dale, 1986) as proposed by Heagerty and Zeger (1996). The (c_1, c_2) global odds ratio describing the association between O_{ij} and O_{ik} is defined as

$$\psi_{i(j,k)(c_1,c_2)} = \frac{\Pr(O_{ij} > c_1, O_{ik} > c_2) \Pr(O_{ij} \leqslant c_1, O_{ik} \leqslant c_2)}{\Pr(O_{ij} > c_1, O_{ik} \leqslant c_2) \Pr(O_{ij} \leqslant c_1, O_{ik} > c_2)}. \tag{2.5}$$

A regression model can be specified for this pairwise association:

$$\log(\psi_{i(j,k)(c_1,c_2)}) = \mathbf{z}^{\mathrm{T}}_{i(j,k)(c_1,c_2)}\boldsymbol{\alpha}, \tag{2.6}$$

$i = 1, 2, \ldots, N,\ j < k = 1, 2, \ldots, n,\ c_1, c_2 = 1, 2, \ldots, C-1$. Here the $\mathbf{z}_{i(j,k)(c_1,c_2)}$ may comprise any relevant covariates for modelling the degree of association between the j and k item responses. Some commonly used examples are

(a) common exchangeable—$\log(\psi_{i(j,k)(c_1,c_2)}) = \alpha_0$,
(b) common item—distance dependent—$\log(\psi_{i(j,k)(c_1,c_2)}) = \alpha_0 + \lambda|j-k|$ (which is particularly relevant when j and k denote times in a longitudinal application) and
(c) full pairwise dependent—$\log(\psi_{i(j,k)(c_1,c_2)}) = \alpha_{c_1c_2} + \lambda_{jk}$.

2.2.2. *Estimation*

Maximum likelihood estimation is one obvious option for estimating parameters; however, the entire likelihood involves many nuisance parameters and is often intractable to specify. When outcome variables are approximately Gaussian, the likelihood method is well developed (Ware, 1985). Few techniques have been available for non-Gaussian outcomes (Stiratelli *et al*., 1984; Fitzmaurice and Laird, 1993). GEEs (Liang and Zeger, 1986) provide a reasonable approach to obtain the parameter estimates under minimal assumptions about the within-cluster dependence for both Gaussian and non-Gaussian outcomes. Analysts may choose from three different GEE methods to estimate the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. First-order GEEs (GEE1—Liang and Zeger (1986)) treat $\boldsymbol{\alpha}$ as a nuisance and focused primarily on obtaining $\boldsymbol{\beta}$. Second-order GEEs (GEE2—Prentice and Zhao (1991)) estimate $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ jointly. Extended alternating logistic regression (ALR—Carey *et al*. (1993)) replaces the estimating equation for $\boldsymbol{\alpha}$ in the GEE1 method by an unbiased non-linear estimating equation and offers high efficiency in the estimation of both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. The standard errors of all three methods are calculated by using robust 'sandwich' variance estimators; details can be found in Heagerty and Zeger (1996). The GEE2 method estimates the association parameters $\boldsymbol{\alpha}$ most precisely; however, it has the disadvantages that the consistency of $\hat{\boldsymbol{\beta}}$ depends on having specified the correct model for the global odds ratios, and that its computational burden quickly grows to infeasibility as data clusters become large. Thus, in situations where inference regarding $\boldsymbol{\beta}$ is primary or when estimation using the GEE2 method is intractable, the GEE1 or ALR approach may be most appropriate.

3. Application

The model described in the previous section has the potential to grow quickly and unreasonably given the possible item and cut-off interactions, and therefore computational intensity is an issue. This section proposes a statistical procedure for efficiently applying the ordinal estimating equation model to multiply measured responses. Broadly, we follow the model building process advocated by McCullagh and Nelder (1989), chapter 12, page 392, which features a loop through model selection and model checking. This loop is continued until we find a statistical model that is appropriate both to the data and to the question to be answered by the analysis. We shall introduce our procedure explicitly in conjunction with the SEE data analysis.

3.1. Model building

3.1.1. Overview

In the SEE project, the focus is on the relationship between reported visual ability and visual impairments adjusting for confounding variables, as described in Section 2.1. Since the far vision subscale comprises substantively unique measurements, item by covariate and cut-off by covariate interactions must be considered. To minimize the computational burden and to achieve a parsimonious modelling procedure, we do the following.

(a) We consult experts to elicit the necessary model complexity, including item by covariate interactions. Where science does not specifically dictate, we perform a preanalysis using standard proportional odds modelling to ensure that substantial item by covariate interactions are accounted for. Backward selection is implemented to settle on a final model. This does not account for within-cluster dependence, but it does estimate the marginal mean parameters consistently; hence, we consider it reasonable for excluding weak predictors (with p-value greater than 0.2) before applying the ordinal estimating equation model.
(b) We fit the ordinal estimating equation model with initial values of marginal mean parameters set as the parameter estimates from the preanalysis. Then we perform an analyst-driven version of backward selection, which selects between included item by covariate interactions according to both scientific and statistical significance.
(c) We diagnose carefully to ensure that the fitted model is not meaningfully misspecified, and we include possible cut-off by covariate interactions.

To fit the ordinal estimating equation model, we must decide on a pairwise association model. We examine the empirical global odds ratio matrix to select a reasonable model for the association structure within clusters, estimating the (c_1, c_2) global odds ratio for the pair (O_{ij}, O_{ik}), $\psi_{i(j,k)(c_1,c_2)}$, empirically from the 2×2 table of the jth *versus* the kth item response dichotomized at levels c_1 and c_2 respectively. In the SEE far vision example, a common global odds ratio assumption $\psi_{i(j,k)(c_1,c_2)} = \psi_{i(j,k)}$ appeared reasonable. Under this assumption, we estimated the empirical log-odds ratio matrix by averaging all the possible (c_1, c_2) log-global-odds ratio estimates for (O_{ij}, O_{ik}). Table 1 shows the empirical log-odds ratio matrix among self-reported visual ability items. These log-odds ratios are substantially and uniformly greater than 0. Because the focus of the SEE project is not on the degree of association between different visual disability items, we chose the common exchangeable model $\log(\psi_{i(j,k)(c_1,c_2)}) = \alpha_0$ as a reasonable and simple association model, and we used the GEE1 method as the estimating method.

Table 1. Empirical log-odds ratios for associations between self-reported visual ability items

Variable	*Log-odds ratios for the following variables:*			
	nightsgn	*daystp*	*dimstp*	*tvwatch*
daysgn	3.93	2.97	2.60	3.37
nightsgn		2.56	2.64	2.67
daystp			3.70	3.06
dimstp				2.85

3.1.2. Initial results

Following the model selection criteria described in the previous subsection, we obtained the initial results shown in Table 2. We shall give a detailed summary of the substantive findings based on our final model (Section 3.3). We focus for now on the technical aspects of interpretation for the multiple-response problem.

(a) Beginning with item effects, the interpretation depends on the levels of interacting variables. For male participants who do not have any comorbid disease and have a visual acuity score equal to 0 (i.e. the familiar '20–20' standard), reading signs at night is the most difficult task, with the lowest odds of reporting better functioning, and watching television is the least difficult task. However, those with really bad visual acuity (e.g. vabnor > 0.72) have less difficulty in walking down steps than watching television and reading street signs.

(b) For most covariates, the effects were deemed comparable across items, supporting the usefulness of reporting summary effects of such covariates on functioning related to distance vision. Several item by covariate interactions identified were

   (i) although women reported more difficulty in far vision items than men did, the discrepancy between genders in levels of difficulty watching television was smaller than in the other far vision activities,
   (ii) there was a negative association between better far vision functioning and the number of comorbid diseases, but this was exaggerated with respect to the walking down steps in daylight and watching television activities and
   (iii) the visual acuity associations varied over items such that those with the steps variables were substantially weaker than with the other far vision items.

(c) The estimated log-global-odds ratio α_0 was 2.04, which indicates strong interitem association but is smaller than in the preliminary analysis. This is because the empirical global odds ratios may include contributions from covariate effects, whereas the association estimated from the ordinal estimating equations model controls for covariates.

3.2. Diagnosis

Having built an initial model, our next step is to assess critically the fit of the model. We now develop graphical displays for this. Graphical diagnostic displays have long proven useful for detecting a lack of model fit to data in ordinary linear regression, and several recent extensions of these displays have been proposed to diagnose generalized linear models (Landwehr *et al.* (1984), Hosmer and Lemeshow (1989), chapter 5, Cook and Weisberg (1997), O'Hara Hines and Carter (1993) and Hall (1995)). Our work emphasizes three methods: cumulative log-odds plots for checking the proportional odds assumption, partial residual plots for investigating possible systematic model departures and Pearson residual plots for evaluating the overall fit. We extend and reformulate these plots from their original application in ordinary linear regression to multiple ordinal measurements.

3.2.1. Cumulative log-odds plot

The ordinal estimating equation model (2.1) assumes that the odds ratios for association between covariates and the event $Y_{ijc} = 1$ are independent of the choice of cut-off c. To check

Table 2. Ordinal estimating equation model for far vision difficulty†

Type	*Variable*	*Initial model estimate*		*Refined model estimate*	
Mean regression					
Intercept	int1	3.5622	(0.5522)	2.9111	(1.0119)
	int2	2.6041	(0.5484)	1.9307	(1.0112)
	int3	1.0777	(0.5485)	0.3979	(1.0131)
Self-reported ability	nightsgn‡	−1.5163	(0.0614)	−2.707	(0.5641)
	daystp	0.7492	(0.1258)	−0.6019	(0.656)
	dimstp	−0.2252	(0.0702)	−2.1318	(0.639)
	tvwatch	0.8777	(0.181)	0.9614	(0.1791)
Confounding variable (main effect)	agec§	−0.0069	(0.0082)	−0.0118	(0.0083)
	mmsec§§	−0.0316	(0.0186)	−0.0306	(0.0187)
	educ*	−0.0275	(0.0134)	−0.0254	(0.0134)
	sex	−0.7515	(0.0837)	−0.7894	(0.0842)
	race	−0.00007	(0.1024)	−0.001	(0.1023)
	comorbid	−0.0956	(0.0246)	−0.119	(0.0344)
	ghqdscor	−0.1754	(0.0484)	−0.1707	(0.0619)
Visual impairment (main effect)	vabnor	−2.8338	(0.309)	−2.2739	(0.3962)
	$(\text{vabnor} - 0.3)^+$	—	—	−1.3547	(0.7374)
	bcs	0.065	(0.0137)	0.0991	(0.0334)
	$(\text{bcs} - 29)^+$	—	—	−0.077	(0.0391)
	difcs	−0.0603	(0.0156)	−0.0689	(0.0433)
	$(\text{difcs} - 2)^+$	—	—	0.0078	(0.0678)
	logster	−0.3381	(0.0836)	−0.3656	(0.0845)
	bestcen	−0.0726	(0.0287)	−0.0751	(0.0287)
Item by covariate interaction	tvwatch × sex	0.3706	(0.1461)	0.3929	(0.1501)
	daystp × comorbid	−0.1006	(0.0329)	−0.1058	(0.0342)
	tvwatch × comorbid	−0.1119	(0.0395)	0.0041	(0.0613)
	dimstp × ghqdscor	—	—	0.0178	(0.0876)
	nightsgn × vabnor	0.6196	(0.2158)	0.4861	(0.29)
	daystp × vabnor	2.3449	(0.3214)	2.619	(0.4118)
	dimstp × vabnor	2.1590	(0.316)	2.4938	(0.3752)
	tvwatch × vabnor	0.614	(0.3078)	—	—
	nightsgn × bcs	—	—	0.0382	(0.0163)
	daystp × bcs	—	—	0.04	(0.0184)
	dimstp × bcs	—	—	0.0566	(0.0184)
Cut-off by (item by) covariate interaction	int3 × comorbid	—	—	0.0235	(0.0273)
	int3 × tvwatch × comorbid	—	—	−0.1453	(0.0472)
	int3 × ghqdscor	—	—	0.036	(0.0473)
	int3 × dimstp × ghqdscor	—	—	−0.1521	(0.0851)
	int3 × nightsgn × $(\text{vabnor} - 0.3)^+$	—	—	1.3486	(0.5086)
	int3 × nightsgn × $(\text{bcs} - 29)^+$	—	—	−0.0216	(0.0113)
	int3 × difcs	—	—	−0.0441	(0.0312)
	int3 × $(\text{difcs} - 2)^+$	—	—	0.1056	(0.0509)
Association regression					
	α_0	2.0419	(0.0903)	2.0492	(0.0902)

†Standard errors are given in parentheses.
‡Reference task, daysgn.
§Age at clinic examination minus 75.
§§Mini-mental state examination score minus 28.
*Years of education minus 11.

this assumption, we compare empirical and model-based relationships between Y_{ijc} and covariates, using cumulative log-odds plots. The cumulative log-odds plot for the jth item and cth level *versus* the kth covariate has y-co-ordinates

$$\log\left\{\frac{\Pr(Y_{ijc}=1)}{1-\Pr(Y_{ijc}=1)}\right\}$$

and x-co-ordinates $(\mathbf{x}_{ij})_k$, where $(\mathbf{x}_{ij})_k$ is the kth element of $\mathbf{x}_{ij}$, $i=1,\ldots,N$. The empirical estimate of the plot is obtained by first smoothing the scatterplot of Y_{ijc} *versus* $(\mathbf{x}_{ij})_k$, then transforming the resulting probability curve to log-odds and finally plotting the transformed curve against $(\mathbf{x}_{ij})_k$. The model-based estimate of the plot is obtained in the same manner, but replacing Y_{ijc} with probabilities $\widehat{\Pr}(Y_{ijc}=1)$ predicted by ordinal estimating equations. By viewing the parallelism of empirical log-odds curves across cutpoints c and comparing empirical with fitted log-odds, we can assess the adequacy of the proportional odds assumption. A similar diagnostic plot was used and proved useful for detecting the lack of model fit in Heagerty and Zeger (1996).

Returning to the far vision example, we first checked for cut-off by item interactions by constructing cumulative log-odds plots with x-co-ordinate equal to the item number (these are not shown). The three empirical cumulative log-odds line segments were parallel to each other and well matched the fitted line segments; thus the marginal item response distributions were well described without cut-off by item interactions. Second, we examined the cumulative log-odds plots against all the model covariates to check the proportional odds assumption. Some of them indicate a lack of fit. For example, Fig. 1(a) displays the item-specific cumulative log-odds plots against the variable bcs. The lack of fit can be described in two ways. Steps variables (daystp, dimstp) and tvwatch have a similar shape, with 'none ($c>3$) *versus* having difficulty', relatively well fitted and log-odds relationships for 'at most little ($c>2$) *versus* at least moderate difficulty' and 'at most moderate ($c>1$) *versus* extreme difficulty' steeper than expected. The curves for both signs variables (daysgn and nightsgn) have a similar shape, with 'at most little *versus* at least moderate difficulty' relatively well fitted and the other two fitted curves gradually apart from the empirical curves. To adjust the violation of the proportional odds assumption, cut-off by covariate, item by covariate and cut-off by item by covariate interactions: (int1 × bcs, int2 × bcs, int3 × bcs), (nightsgn × bcs, int1 × nightsgn × bcs, int3 × nightsgn × bcs), (daystp × bcs, int1 × daystp × bcs, int2 × daystp × bcs), (dimstp × bcs, int1 × dimstp × bcs, int2 × dimstp × bcs) and (tvwatch × bcs, int1 × tvwatch × bcs, int2 × tvwatch × bcs) may be needed.

3.2.2. *Partial residual plot*

Partial residual plots can help to assess and model substantial non-linearity in relationships between outcomes and predictor variables. Consider the model

$$\begin{aligned} E(\mathbf{y}) &= \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta} + f(\mathbf{z}), \\ \operatorname{cov}(\mathbf{y}) &= \sigma^2 I, \end{aligned} \tag{3.1}$$

where $f(\mathbf{z})$ is some function of a covariate $\mathbf{z}$. For linear $f(\mathbf{z}) = \mathrm{z}^{\mathrm{T}}\alpha$, Larsen and McCleary (1972) defined the partial residual vector as

$$\mathbf{r}^* = \mathbf{r}_{\mathbf{y}|\mathbf{X},\mathbf{z}} + \mathbf{z}^{\mathrm{T}}\hat{\alpha}, \tag{3.2}$$

where $\mathbf{r}_{\mathbf{y}|\mathbf{X},\mathbf{z}} = \mathbf{y} - \mathbf{X}^{\mathrm{T}}\hat{\boldsymbol{\beta}} - \mathbf{z}^{\mathrm{T}}\hat{\alpha}$. Here, $\hat{\boldsymbol{\beta}}$ and $\hat{\alpha}$ are the least squares estimates obtained by fitting the complete model (3.1). It is easy to show that $E(\mathbf{r}^*) = \mathbf{z}^{\mathrm{T}}\alpha$. So, we would expect the plot of

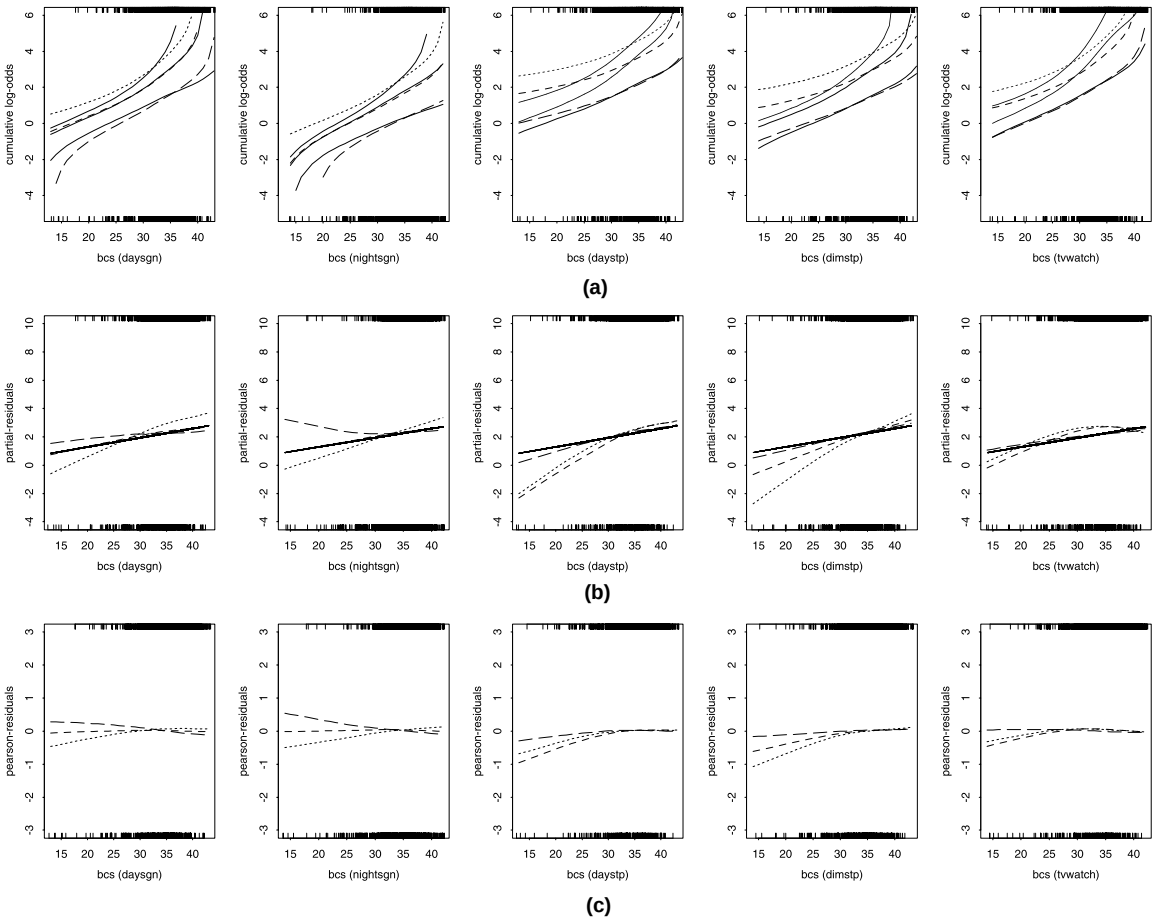

Fig. 1. Diagnostic plots for contrast sensitivity scores (bcs), under the initial model: (a) cumulative log-odds plots (———, empirical log-odds; ·····, fitted log-odds from the marginal mean model, $c > 1$; – – – –, fitted log-odds, $c > 2$; — — —, fitted log-odds, $c > 3$); (b) partial residual plots (———, estimated linear relationship from our model; ·····, cutpoint > 1; – – – –, cutpoint > 2; — — —, cutpoint > 3); (c) Pearson residual plots (·····, cutpoint > 1; – – – –, cutpoint > 2; — — —, cutpoint > 3); tick marks at the top and bottom of each box indicate $(\mathbf{x}_{ij})_k$-values for people who did not have any difficulty doing the jth item activity ($Y_{ij3} = 1$) and for people who had difficulty doing the jth item activity ($Y_{ij3} = 0$) respectively

$\mathbf{r}^*$ *versus* $\mathbf{z}$ to be linear. When $f(\mathbf{z})$ is non-linear, $E(\mathbf{r}^*)$ is generally not exactly $f(\mathbf{z})$, and an accurate determination of the structure of f can be difficult. However, the partial residual plot reasonably suggests, and thus is useful to determine, the form of f.

Landwehr *et al.* (1984) extended the partial residual plot to logistic regression. Consider the model $\text{logit}(\mathbf{p}) = \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta} + f(\mathbf{z})$, where $\mathbf{p} = \Pr(\mathbf{y} = 1)$. If $f(\mathbf{z}) = \mathbf{z}^{\mathrm{T}}\boldsymbol{\alpha}$, then

$$\text{logit}(\mathbf{p}) = \boldsymbol{\eta} = \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta} + \mathbf{z}^{\mathrm{T}}\alpha = \mathbf{X}^{*\mathrm{T}}\boldsymbol{\beta}^*,$$

where $\mathbf{X}^* = (\mathbf{X}^{\mathrm{T}}, \mathbf{z}^{\mathrm{T}})^{\mathrm{T}}$ and $\boldsymbol{\beta}^* = (\boldsymbol{\beta}^{\mathrm{T}}, \alpha)^{\mathrm{T}}$. Typically $\boldsymbol{\beta}^*$ is estimated by the method of iteratively reweighted least squares, so $\hat{\boldsymbol{\beta}}^*$ at the $(t + 1)$th iteration can be written (O'Hara Hines and Carter, 1993)

$$\hat{\boldsymbol{\beta}}^*(t+1) = (\mathbf{X}^*\hat{W}(t)\mathbf{X}^{*\mathrm{T}})^{-1}(\mathbf{X}^*\hat{W}(t)\mathbf{y}^*(t)); \tag{3.3}$$

here,

$$W = \left(\frac{\partial \mathbf{p}}{\partial \boldsymbol{\eta}}\right)\text{cov}(\mathbf{y})^{-1}\left(\frac{\partial \mathbf{p}}{\partial \boldsymbol{\eta}}\right),$$

$\hat{W}(t)$ refers to the value of W evaluated at $\hat{\boldsymbol{\beta}}^*(t)$, $\hat{\mathbf{p}}(t)$ is the estimate of $\mathbf{p}$ at the tth iteration and

$$\mathbf{y}^*(t) = \mathbf{X}^{*\mathrm{T}}\hat{\boldsymbol{\beta}}^*(t) + \left(\frac{\partial \boldsymbol{\eta}}{\partial \mathbf{p}}\right)(\mathbf{y} - \hat{\mathbf{p}}(t)).$$

Comparing with weighted linear regression, $\mathbf{y}^*$, $\mathbf{X}^{*\mathrm{T}}$, $\hat{\boldsymbol{\beta}}^*$ and $(\partial\boldsymbol{\eta}/\partial\mathbf{p})(\mathbf{y} - \hat{\mathbf{p}})$ can be thought of as observation, fitted value and residual. So, using the fact that the logit link is the canonical link function for binomial data (thus, $(\partial\boldsymbol{\eta}/\partial\mathbf{p}) = \widehat{\text{cov}}(\mathbf{y})^{-1}$), the logistic partial residual is

$$\hat{\mathbf{r}}^*_{\text{log}} = \widehat{\text{cov}}(\mathbf{y})^{-1}(\mathbf{y} - \hat{\mathbf{p}}) + \mathbf{z}^{\mathrm{T}}\hat{\alpha}. \tag{3.4}$$

We proceed to develop partial residuals for the ordinal estimating equation model: $\text{logit}\{E(Y_{ijc})\} = \theta_c + \mathbf{x}_{ij}^{\mathrm{T}}\boldsymbol{\beta} + f(z_{ij})$. Let

$$\mathbf{Y}_i = (Y_{i11}, Y_{i12}, \ldots, Y_{i1(C-1)}, Y_{i21}, Y_{i22}, \ldots, Y_{i2(C-1)}, \ldots, Y_{in1}, Y_{in2}, \ldots, Y_{in(C-1)})^{\mathrm{T}},$$

$$\mathbf{Y} = (\mathbf{Y}_1^{\mathrm{T}}, \mathbf{Y}_2^{\mathrm{T}}, \ldots, \mathbf{Y}_N^{\mathrm{T}})^{\mathrm{T}},$$

$$p_{ijc} = \Pr(Y_{ijc} = 1),$$

$$\mathbf{p}_i = \Pr(\mathbf{Y}_i = 1),$$

$$\mathbf{p}_{\text{ord}} = \Pr(\mathbf{Y} = 1),$$

$$\boldsymbol{\eta}_{\text{ord}} = \text{logit}(\mathbf{p}_{\text{ord}}).$$

The estimating equation for parameters in model (2.1) can be shown to have similar iteratively reweighted least squares equation (3.3) (see Appendix A for details). If $f(z_{ij}) = z_{ij}\alpha$, using reasoning analogous to that for the logistic partial residual yields the partial residual of the ordinal estimating equation model as

$$\hat{\mathbf{r}}^*_{\text{ord}} = \left(\frac{\partial \boldsymbol{\eta}_{\text{ord}}}{\partial \mathbf{p}_{\text{ord}}}\right)(\mathbf{Y} - \hat{\mathbf{p}}_{\text{ord}}) + \mathbf{z}^{\mathrm{T}}\hat{\alpha}, \tag{3.5}$$

where $(\partial\boldsymbol{\eta}_{\text{ord}}/\partial\mathbf{p}_{\text{ord}})$ is an $Nn(C-1) \times Nn(C-1)$ diagonal matrix with $((i-1)n(C-1) + (j-1)(C-1) + c)$th element

$$\frac{\mathrm{d}\{\mathrm{logit}(p_{ijc})\}}{\mathrm{d}p_{ijc}} = \frac{1}{p_{ijc}(1-p_{ijc})}$$

and $\mathbf{z}$ is the extended vector of z_{ij} with respect to $\mathbf{Y}$. The partial residual plot for the jth item and cth level *versus* $\mathbf{z}$ is the scatterplot of $(\hat{\mathbf{r}}^{*}_{\mathrm{ord}})_{ijc}$ (y-co-ordinate) *versus* z_{ij} (x-co-ordinate) over all i.

Our plot is a direct extension of the logistic partial residual plot. Therefore, according to the findings of Landwehr *et al*. (1984), the shape of the proposed partial residual plot for multiple ordinal measurements can reasonably suggest the relationship between the log-odds of Y_{ijc} and z_{ij}. We also plot the estimated linear relationship from our model with y-co-ordinate $\mathbf{z}^{\mathrm{T}}\hat{\alpha}$ and x-co-ordinate $\mathbf{z}$, and we use cubic smoothing splines to help to discern average patterns.

Fig. 1(b) displays partial residuals *versus* bcs for the five items. Summarizing the information from these plots, we found that the relationship between bcs and self-reported visual functioning differed noticeably on the bcs ranges 0–29 *versus* 29 and above, especially at the third level of nightsgn. To address this, we added linear spline terms $(\mathrm{bcs} - 29)^{+}$ and $\mathrm{int3} \times \mathrm{nightsgn} \times (\mathrm{bcs} - 29)^{+}$ to the model, where $X^{+} = X$ if $X > 0$ and $X^{+} = 0$ if $X \leqslant 0$.

3.2.3. *Pearson residual plot*

The residual *versus* covariate plot is the most often used graphical method for assessing the goodness of fit in ordinary linear regression. Analogously, Pearson residuals are often used with polytomous data. Following the definition in McCullagh and Nelder (1989), page 37, the Pearson residual for Y_{ijc} in model (2.1) is defined as

$$e_{ijc} = \frac{Y_{ijc} - \hat{p}_{ijc}}{\{\hat{p}_{ijc}(1-\hat{p}_{ijc})\}^{1/2}},$$

where $\hat{p}_{ijc}$ is the estimate of p_{ijc} with θ_c and $\boldsymbol{\beta}$ replaced by $\hat{\theta}_c$ and $\hat{\boldsymbol{\beta}}$ respectively. The Pearson residual plot for the jth item and cth level *versus* the kth covariate can be obtained by drawing e_{ijc} (the y-co-ordinate) *versus* $(\mathbf{x}_{ij})_k$ (the x-co-ordinate) over all i. The same smoothing method is used as for the partial residual plot. If the specified model is correct, we shall have a plot with slope and intercept near 0. Any systematic pattern might suggest a possible lack of fit (McCullagh and Nelder (1989), page 399).

Fig. 1(c) shows the Pearson residuals for all three levels *versus* bcs on five different items. The curves for both steps items gradually separate from 0 as the values of bcs become smaller (they have overestimated the probability of reporting better visual functioning for small bcs). This indicates a need for daystp × bcs and dimstp × bcs terms. The systematic break point at 29 for the third level of nightsgn can be seen. These findings are consistent with those from partial residual plots.

3.2.4. *Influential points*

In smoothing our plots, leverage points and sparse data can have a substantial effect on the appearance of the curves. To mitigate this problem, we first eliminated clearly isolated points before smoothing, and then smoothed along the middle 99% of the 'x'-variable in each plot. For example, we identified two extreme leverage points, representing people with a visual acuity greater than 1.5 (almost blind) who reported perfect ability in doing both steps tasks. Although these reports may be valid, they dramatically contradicted the reporting tendency

among people who had a visual acuity less than 1.5. To avoid obscuring the relationship for vabnor < 1.5, we removed the results for these two people from the plots and also from the refined models.

3.3. Result for the refined model

3.3.1. Refined model

Using the diagnostic findings, we refitted the ordinal estimating equation model. Table 2 shows the refined ordinal estimating equation model. The model is non-hierarchical. Two three-way interactions, int3 $\times$ nightsgn $\times$ $(\text{vabnor} - 0.3)^+$ and int3 $\times$ nightsgn $\times$ $(\text{bcs} - 29)^+$, are specified without their corresponding two-way interactions, which implies that the increased slope for vabnor > 0.3, for example, is the same across items and difficulty levels, except that the association is further enlarged in the tvwatch item for comparing none *versus* having visual difficulty. In the SEE study, interactions with spline terms are not the major interest. However, diagnostic plots (which are not shown here) indicated an apparent lack of fit without the above three-way interactions. Also a model with all the corresponding two-way interactions did not converge owing to the sparseness of data. We therefore decided to add them alone.

Although the new model corrected much of the bias that we previously identified, some of the plots still indicated a lack of fit. For example, in the partial residuals plot (Fig. 2(b)), we can still find a curvilinear relationship between self-reported visual ability and bcs at the third level of nightsgn after adjusting for the cut-off by covariate interaction and spline term. This might suggest a higher order spline term. However, the Pearson residuals plot (Fig. 2(b)) did not indicate a poor fit in this term. Since the high order relationship between reported visual ability and bcs was not our major interest, to avoid the possible additional complexity we decided not to correct this.

3.3.2. Substantive findings

The ordinal estimating equations approach illuminated the following scientific questions.

(a) Even after controlling for vision, *women* reported more difficulty in performing the far vision activities than *men* did. Also, the higher the number of years of education, the number of comorbid diseases and the general health questionnaire depression score, the more difficulty was reported with the far vision activities. These findings persisted across different visual difficulties, but in many cases the strengths of associations varied across difficulties. The above findings suggest that it may be necessary to tailor the assessment and treatment of visual disability to characteristics other than vision. Interestingly, age was not associated with far vision functioning after adjusting for visual impairment.

(b) All vision predictors independently and non-trivially predicted visual functioning after controlling for confounders. This finding contradicts the current clinical practice, which focuses primarily on visual acuity. This is an important contribution from the SEE study.

(c) Several differential associations across vision items were found. These would have been missed by scoring analysis. Most importantly, the associations between visual acuity and the steps variables were substantially weaker than with the other far visual functioning, and the association between contrast sensitivity and the steps variables were substantially stronger than with the other far vision variables. This suggests

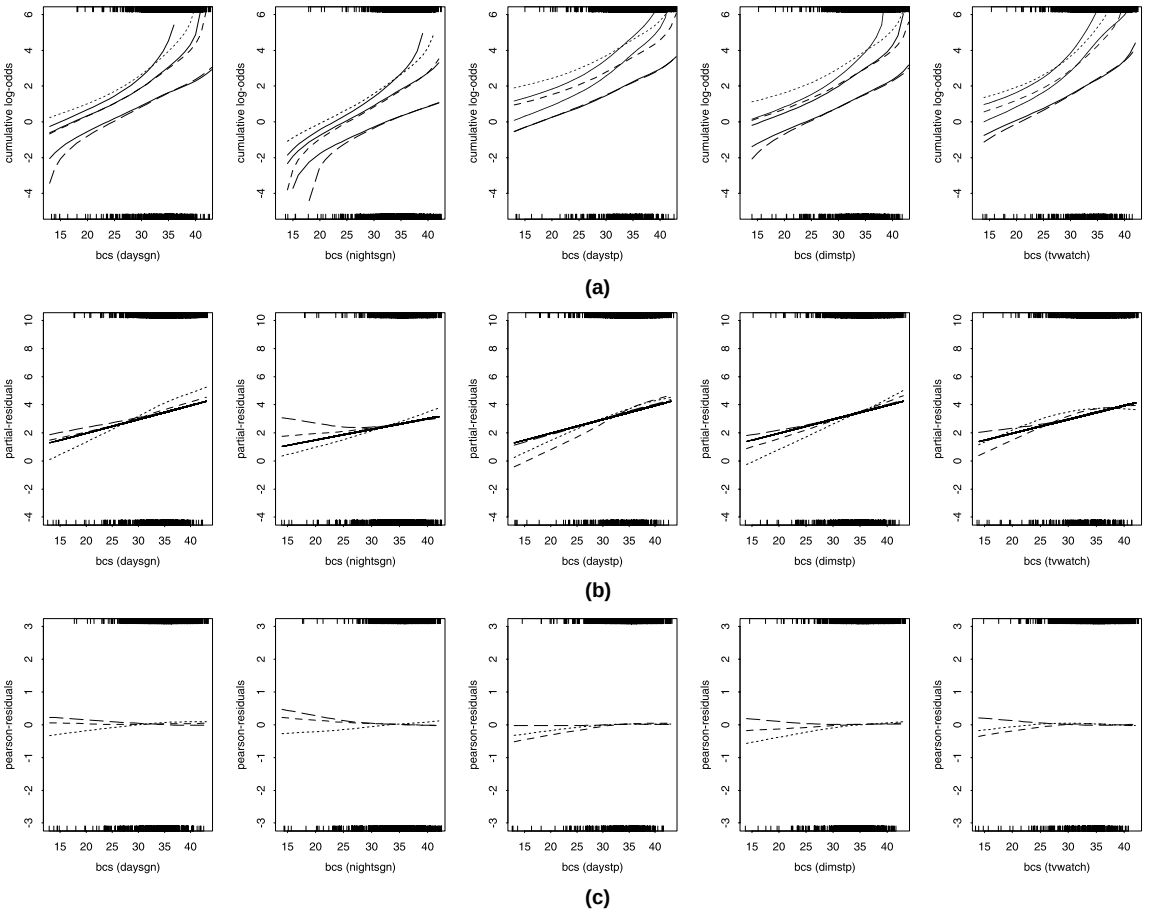

Fig. 2. Diagnostic plots for contrast sensitivity scores (bcs), under the refined model: (a) cumulative log-odds plots (———, empirical log-odds; ·····, fitted log-odds from the marginal mean model, $c > 1$; – – – –, fitted log-odds, $c > 2$; — — —, fitted log-odds, $c > 3$); (b) partial residual plots (———, estimated linear relationship from our model; ·····, cutpoint > 1; – – – –, cutpoint > 2; — — —, cutpoint > 3); (c) Pearson residual plots (·····, cutpoint > 1; – – – –, cutpoint > 2; — — —, cutpoint > 3); tick marks at the top and bottom of each box indicate $(\mathbf{x}_{ij})_k$-values for people who did not have any difficulty doing the jth item activity ($Y_{ij3} = 1$) and for people who had difficulty doing the jth item activity ($Y_{ij3} = 0$) respectively

distinct resolution and contrast components of distance vision, which is consistent with Rubin *et al*. (1994), and highlights the possibility that assessing non-acuity impairments might prevent falls and other adverse outcomes.

(d) Curvilinear relationships were found in visual acuity, contrast sensitivity and glare sensitivity. The estimated visual acuity association with far vision showed that elders with acuity worse than 0.3 were particularly at risk for poor far vision function. In contrast sensitivity, there was a threshold (equalling 29) above which contrast sensitivity declines had little effect, but below which there was a large effect.

4. Evaluation

In this section, we evaluate the benefit of incorporating interitem associations in fitting the ordinal estimating equation model and the trade-off between simple *versus* complex models. We address the benefits of acknowledgeing and estimating interitem associations first. The ordinal estimating equation model allows us to specify associations within clusters. The consistency of the $\boldsymbol{\beta}$-estimators obtained from the GEE1 or ALR methods does not require a correct specification of the association model, but using grossly incorrect working association structures can lead to inefficiency of the estimator (Liang *et al*., 1992; Fitzmaurice and Laird, 1993).

In the following, we use the SEE far vision example to compare various methods of estimating the proportional odds model. Table 3 summarizes three different estimating methods:

(a) 'naïve' proportional odds fitting ignoring item associations both in fitting and in computing standard errors;
(b) ordinal estimating equation fitting with the GEE1 method and independent item working association, but correcting standard errors by using robust sandwich estimators;
(c) ordinal estimating equation fitting with the GEE1 method and exchangeable item association.

The naïve model (a) was obtained by the SAS procedure `LOGISTIC`. Since this procedure cannot fit cut-off by covariate interactions, we replaced all cut-off by item by covariate interactions in the final model with their associated item by covariate interactions, and then fitted a model with those and remaining terms. Because the existing ordinal estimating equation software cannot be implemented with independent working covariance (i.e. $\boldsymbol{\alpha} = 0$), the parameter estimates of model (b) were obtained by using the $\boldsymbol{\beta}$-estimates from the naïve model, and the corrected standard errors for the independent association model were estimated by using the robust sandwich variance estimator $\mathbf{AV}_{\text{gee1}}$ on page 1029 of Heagerty and Zeger (1996), with $\mathbf{V}_{i11}$ being defined by expressions (a) and (b) in Appendix A and expression (c) there being redefined as 0. All three methods produced similar parameter estimates because of the consistency property of $\hat{\boldsymbol{\beta}}$ under the GEE1 method. Comparing the coefficient standard errors between ordinal estimating equations with independent association and ordinal estimating equations with exchangeable association, the exchangeable model standard errors were 9% lower on average than the independence model standard errors. This supports the notion that the item associations are non-ignorable in the SEE study. The naïve standard error estimates varied widely because they do not correct standard errors to account for association in the estimating equation, and hence the standard error estimates are invalid.

Table 3. Comparison of the proportional odds model and ordinal estimating equation model†

Type	*Variable*	*Proportional odds model estimate*		*Ordinal estimating equation model (independent) estimate*		*Ordinal estimating equation model (exchangeable) estimate*	
Mean regression							
Intercept	int1	3.0937	(0.6332)	3.0937	(0.9693)	3.1257	(0.9656)
	int2	2.1322	(0.6329)	2.1322	(0.967)	2.1538	(0.965)
	int3	0.6028	(0.6327)	0.6028	(0.969)	0.7199	(0.968)
Self-reported ability	nightsgn‡	−1.839	(1.3562)	−1.839	(1.0815)	−1.8252	(0.7641)
	daystp	−0.5144	(0.8246)	−0.5144	(0.6842)	−0.7917	(0.6492)
	dimstp	−1.7509	(0.7533)	−1.7509	(0.6598)	−2.1705	(0.618)
	tvwatch	0.8782	(0.1996)	0.8782	(0.1773)	0.9585	(0.1776)
Confounding variable (main effect)	agec§	−0.01	(0.0058)	−0.01	(0.0085)	−0.0112	(0.0083)
	mmsec§§	−0.0203	(0.0122)	−0.0203	(0.0193)	−0.0302	(0.0186)
	educ*	−0.023	(0.009)	−0.023	(0.0136)	−0.026	(0.0134)
	sex	−0.7881	(0.0613)	−0.7881	(0.0874)	−0.7764	(0.084)
	race	0.0926	(0.0694)	0.0926	(0.1079)	−0.0033	(0.1023)
	comorbid	−0.0946	(0.0196)	−0.0946	(0.0254)	−0.0977	(0.0246)
	ghqdscor	−0.1651	(0.034)	−0.1651	(0.0473)	−0.1511	(0.0502)
Visual impairment (main effect)	vabnor	−2.5899	(0.3374)	−2.5899	(0.4396)	−2.6031	(0.4189)
	$(\text{vabnor} - 0.3)^+$	−0.3932	(0.4849)	−0.3932	(0.784)	−0.4927	(0.7699)
	bcs	0.0959	(0.0202)	0.0959	(0.0321)	0.0824	(0.0325)
	$(\text{bcs} - 29)^+$	−0.0762	(0.0237)	−0.0762	(0.0386)	−0.0659	(0.0385)
	difcs	−0.1079	(0.0205)	−0.1079	(0.0332)	−0.1076	(0.0298)
	$(\text{difcs} - 2)^+$	0.0944	(0.0327)	0.0944	(0.0513)	0.0934	(0.048)
	logster	−0.3965	(0.0578)	−0.3965	(0.0848)	−0.3717	(0.0841)
	bestcen	−0.0966	(0.0182)	−0.0966	(0.0302)	−0.0762	(0.0286)
Item by covariate interaction	tvwatch × sex	0.4807	(0.1647)	0.4807	(0.1475)	0.3752	(0.1485)
	daystp × comorbid	−0.0956	(0.0458)	−0.0956	(0.0361)	−0.1015	(0.0332)
	tvwatch × comorbid	−0.1254	(0.0473)	−0.1254	(0.0411)	−0.1163	(0.0412)
	dimstp × ghqdscor	−0.114	(0.072)	−0.114	(0.0573)	−0.0933	(0.0522)
	nightsgn × vabnor	1.193	(0.5328)	1.193	(0.4644)	1.0741	(0.4155)
	daystp × vabnor	2.4712	(0.4684)	2.4712	(0.4231)	2.6733	(0.4166)
	dimstp × vabnor	2.5139	(0.4328)	2.5139	(0.4075)	2.6385	(0.3825)
	nightsgn × $(\text{vabnor} - 0.3)^+$	−2.2408	(1.1531)	−2.2408	(1.0817)	−1.9561	(0.8741)
	nightsgn × bcs	0.0064	(0.0471)	0.0064	(0.0381)	0.0046	(0.0265)
	daystp × bcs	0.037	(0.0232)	0.037	(0.0192)	0.0452	(0.0182)
	dimstp × bcs	0.0461	(0.0215)	0.0461	(0.0189)	0.0575	(0.0177)
	nightsgn × $(\text{bcs} - 29)^+$	0.0216	(0.0549)	0.0216	(0.0469)	0.031	(0.0341)
Association regression							
	α_0	—	—	—	—	2.054	(0.0912)

†Standard errors are given in parentheses.
‡Reference task, daysgn.
§Age at clinic examination minus 75.
§§Mini-mental state examination score minus 28.
*Years of education minus 11.

Next, we shall discuss the trade-off between complex and simple mean models. The descriptive specificity of our approach is arguably both a strength and a weakness. It may illuminate scientific questions that are not addressed by simpler analytic approaches. However, our approach risks overfitting data or obscuring the central findings with trivial detail if it is applied haphazardly. In primarily inferential analyses, science must strongly dictate the areas in which model refinements are substantively necessary or important;

arguably, other refinements should be made only if they indicate a severe lack of model fit. For instance, the science of the SEE analysis identified item by vision interactions as very important. Curvilinear relationships between vision measures and visual functioning may suggest thresholds that distinguish declines as having minor *versus* major effects on functioning and so are also of considerable importance. In contrast, cutpoint by item by vision interactions have much lesser importance and merit being added to the model if the analysis is primarily descriptive and the fit is poor. Despite the interesting suggestion of different relationships of bcs with less severe *versus* severe difficulty of negotiating signs in daylight and dim steps (Fig. 2(b)), we opted not to refine the model because there was no evidence showing a serious lack of fit (Fig. 2(c)). However, the interactions int3 × nightsgn × $(\text{vabnor} - 0.3)^+$ and int3 × nightsgn × $(\text{bcs} - 29)^+$ were included since diagnostic plots (which are not shown here) indicated an apparent poor fit without them.

5. Discussion

Clustered ordinal outcomes are very common in biomedical, social and behavioural science studies. Determining the dependence of such responses on covariates remains an analytic challenge. In this paper, we applied the ordinal estimating equations approach (Heagerty and Zeger, 1996) to data from the SEE project. In doing so, we highlighted the complexities of analysing distinct, as opposed to identical, item responses, and we proposed modelling to accommodate such responses. We also modified the cumulative log-odds and Pearson residual plots, and reformulated the partial residual plot to evaluate a lack of model fit. The result is a labour-intensive but workable strategy for assuring a reasonable descriptive fit of the ordinal estimating equation model to multiple ordinal responses.

Simple generalizations of the HZ model can increase the flexibility of the models proposed. HZ models allow subjects to have different numbers of reported difficulty items, n_i. Thus, the situation of non-responses for some participants subject to missingness completely at random (Little and Rubin, 1987) can be handled in the models proposed by adding dummy variables indicating different items, and only including answered items and their associated covariates. Also, this model may be generalized to allow a different number of categories, C_j, per item. When the numbers vary, we must allow item and cut-off specific intercepts θ_{jc} and item-specific $\boldsymbol{\beta}_j$ coefficients per equation (2.4) to state correctly the proportional odds assumption within item. This is because now categories in each item might represent different meanings.

In the refined model, we added several cut-off by covariate interactions to correct potential violations of the proportional odds assumption. However, the usefulness of adding interactions to allow non-parallel regression lines across cutpoints is limited by the fact that the lines must intersect, and the estimated category probabilities $\widehat{\Pr}(O_{ij} > c)$ could be negative for some values of covariates (McCullagh and Nelder (1989), page 155). The cumulative log-odds plots after adding these cutpoint interactions (e.g. Fig. 2(a)) show no intersections between fitted lines within the range of covariate space. Thus, the potential for negative fitted probabilities is not serious.

The software for fitting the HZ model can be downloaded from the Web site

`http://www.jhsph.edu/biostats/software.html`

under the category 'Estimating equations for dependent ordinal data'. An S-PLUS function that implements the graphical diagnostic methods proposed is available from the authors.

Acknowledgements

This work was supported by National Institute on Aging program project P01-AG-10184-03 and National Institutes of Mental Health grant 5 R01 MH 56639. Dr Bandeen-Roche is a Brookdale National Fellow. The authors wish to thank the Joint Editor and two referees for their valuable comments, which led to a great improvement in this paper.

Appendix A: Generalized residuals for the ordinal estimating equation model

The estimating equation for parameters in model (2.1) using the GEE1 method is as follows (Heagerty and Zeger (1996), bottom left-hand side of page 1029):

$$\hat{\boldsymbol{\beta}}_{\text{ord}}(t+1) = \hat{\boldsymbol{\beta}}_{\text{ord}}(t) + \left(\sum_{i=1}^{N} \mathbf{D}_{i11}^{\mathrm{T}} \hat{\mathbf{V}}_{i11}^{-1} \mathbf{D}_{i11}\right)^{-1} \sum_{i=1}^{N} \mathbf{D}_{i11}^{\mathrm{T}} \hat{\mathbf{V}}_{i11}^{-1} \{\mathbf{Y}_i - \hat{\mathbf{p}}_i(t)\}, \tag{A.1}$$

where $\hat{\boldsymbol{\beta}}_{\text{ord}}(t)$ is the tth iteration estimate of $\boldsymbol{\beta}_{\text{ord}} = (\theta_1, \ldots, \theta_{C-1}, \boldsymbol{\beta}^{\mathrm{T}})^{\mathrm{T}}$, $\mathbf{D}_{i11} = (\partial \mathbf{p}_i / \partial \boldsymbol{\beta}_{\text{ord}})$ and $\mathbf{V}_{i11} = \text{cov}(\mathbf{Y}_i)$. Notice that

$$\mathbf{D}_{i11} = \left(\frac{\partial \mathbf{p}_i}{\partial \boldsymbol{\beta}_{\text{ord}}}\right) = \left(\frac{\partial \mathbf{p}_i}{\partial \boldsymbol{\eta}_{i,\text{ord}}}\right)\left(\frac{\partial \boldsymbol{\eta}_{i,\text{ord}}}{\partial \boldsymbol{\beta}_{\text{ord}}}\right) = \left(\frac{\partial \mathbf{p}_i}{\partial \boldsymbol{\eta}_{i,\text{ord}}}\right)\mathbf{X}_{i,\text{ord}}^{\mathrm{T}}, \tag{A.2}$$

where $\boldsymbol{\eta}_{i,\text{ord}} = \text{logit}(\mathbf{p}_i)$ and $\mathbf{X}_{i,\text{ord}}$ is the design matrix of person i based on model (2.1). Using model (2.2) as an example to illustrate the design matrix, person i's marginal mean can be written as the product of $\mathbf{X}_{i,\text{ord}}$ and $\boldsymbol{\beta}_{\text{ord}}$:

$$\text{logit}(\mathbf{p}_i) = \text{logit}\begin{pmatrix} E(Y_{i11}) \\ E(Y_{i12}) \\ E(Y_{i13}) \\ E(Y_{i21}) \\ E(Y_{i22}) \\ E(Y_{i23}) \\ \vdots \\ E(Y_{i51}) \\ E(Y_{i52}) \\ E(Y_{i53}) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & \text{sex}_i \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & \text{sex}_i \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & \text{sex}_i \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & \text{sex}_i \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & \text{sex}_i \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & \text{sex}_i \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & \text{sex}_i \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & \text{sex}_i \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & \text{sex}_i \end{pmatrix} \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \\ \gamma_5 \\ \beta \end{pmatrix} = \mathbf{X}_{i,\text{ord}}^{\mathrm{T}} \boldsymbol{\beta}_{\text{ord}}.$$

The various elements of $\mathbf{V}_{i11} = (\text{cov}(Y_{ij_1c_1}, Y_{ij_2c_2}))$ are as follows:

(a) $j_1 = j_2 = j$, $c_1 = c_2 = c$,

$$\text{cov}(Y_{ij_1c_1}, Y_{ij_2c_2}) = \text{var}(Y_{ijc}) = p_{ijc}(1 - p_{ijc};$$

(b) $j_1 = j_2 = j$, $c_1 > c_2$,

$$\begin{aligned} \text{cov}(Y_{ij_1c_1}, Y_{ij_2c_2}) &= \Pr(Y_{ijc_1} = 1, Y_{ijc_2} = 1) - p_{ijc_1}p_{ijc_2} \\ &= \Pr(Y_{ijc_1} = 1) - p_{ijc_1}p_{ijc_2} = p_{ijc_1}(1 - p_{ijc_2}); \end{aligned}$$

(c) $j_1 \neq j_2$,

$$\text{cov}(Y_{ij_1c_1}, Y_{ij_2c_2}) = S_{j_1j_2} - p_{ij_1c_1}p_{ij_2c_2},$$

where $S_{j_1j_2} = \Pr(Y_{ij_1c_1} = 1, Y_{ij_2c_2} = 1) = \Pr(O_{ij_1} > c_1, O_{ij_2} > c_2)$. To derive a useful closed form for $S_{j_1j_2}$, notice from equation (2.5) that

$$\psi_{i(j_1,j_2)(c_1,c_2)} = \frac{S_{j_1j_2}(1 - p_{ij_1c_1} - p_{ij_2c_2} + S_{j_1j_2})}{(p_{ij_1c_1} - S_{j_1j_2})(p_{ij_2c_2} - S_{j_1j_2})}.$$

Thus,

$$S_{j_1 j_2} = \begin{cases} \frac{1}{2}(\psi_{i(j_1,j_2)(c_1,c_2)} - 1)^{-1}\{1 + (p_{ij_1c_1} + p_{ij_2c_2})(\psi_{i(j_1,j_2)(c_1,c_2)} - 1) \\ -G(p_{ij_1c_1},\ p_{ij_2c_2},\ \psi_{i(j_1,j_2)(c_1,c_2)})\} & \text{if } \psi_{i(j_1,j_2)(c_1,c_2)} \neq 1, \\ p_{ij_1c_1}p_{ij_2c_2} & \text{if } \psi_{i(j_1,j_2)(c_1,c_2)} = 1, \end{cases}$$

where

$$G(p_{ij_1c_1}, p_{ij_2c_2}, \psi_{i(j_1,j_2)(c_1,c_2)}) = [\{1 + (p_{ij_1c_1} + p_{ij_2c_2})(\psi_{i(j_1,j_2)(c_1,c_2)} - 1)\}^2 - 4\psi_{i(j_1,j_2)(c_1,c_2)}(\psi_{i(j_1,j_2)(c_1,c_2)} - 1)p_{ij_1c_1}p_{ij_2c_2}]^{1/2}.$$

By substituting equation (A.2) into equation (A.1), we can then obtain

$$\hat{\boldsymbol{\beta}}_{\text{ord}}(t+1) = (\mathbf{X}_{\text{ord}}\hat{W}_{\text{ord}}(t)\mathbf{X}_{\text{ord}}^{\mathrm{T}})^{-1}(\mathbf{X}_{\text{ord}}\hat{W}_{\text{ord}}(t)\mathbf{Y}^*(t)), \tag{A.3}$$

where $\mathbf{X}_{\text{ord}} = (\mathrm{X}_{1,\text{ord}}, \ldots, \mathbf{X}_{N,\text{ord}})$,

$$\hat{W}_{\text{ord}}(t) = \left(\frac{\partial \mathbf{p}_{\text{ord}}}{\partial \boldsymbol{\eta}_{\text{ord}}}\right)\widehat{\text{cov}}(\mathbf{Y})^{-1}\left(\frac{\partial \mathbf{p}_{\text{ord}}}{\partial \boldsymbol{\eta}_{\text{ord}}}\right)$$

and

$$\mathbf{Y}^*(t) = \mathbf{X}_{\text{ord}}^{\mathrm{T}}\hat{\boldsymbol{\beta}}_{\text{ord}}(t) + \left(\frac{\partial \boldsymbol{\eta}_{\text{ord}}}{\partial \mathbf{p}_{\text{ord}}}\right)(\mathbf{Y} - \hat{\mathbf{p}}_{\text{ord}}(t)).$$

Notice that equation (A.3) has the same iteratively reweighted least squares equation as equation (3.3).

References

Bailey, I. L., Bullimore, M. A., Raasch, T. W. and Taylor, H. R. (1991) Clinical grading and the effects of scaling. *Invest. Ophth. Vis. Sci.*, **32**, 422–432.

Bandeen-Roche, K., Huang, G. H., Munoz, B. and Rubin, G. S. (1999) On determining risk factor associations with questionnaire outcomes: a methods case study. *Am. J. Epidem.*, **150**, 1165–1178.

Bandeen-Roche, K., Miglioretti, D. L., Zeger, S. L. and Rathouz, P. J. (1997) Latent variable regression for multiple outcomes. *J. Am. Statist. Ass.*, **92**, 1375–1386.

Bartholomew, D. J. (1987) *Latent Variable Models and Factor Analysis*. London: Griffin.

Bryk, A. S. and Raudenbush, S. W. (1992) *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park: Sage.

Carey, V. J., Zeger, S. L. and Diggle, P. (1993) Modelling multivariate binary data with logistic regressions. *Biometrika*, **80**, 517–526.

Cook, R. D. and Weisberg, S. (1997) Graphics for assessing the adequacy of regression models. *J. Am. Statist. Ass.*, **92**, 490–499.

Dale, J. R. (1986) Global cross-ratio models for bivariate, discrete, ordered responses. *Biometrics*, **42**, 909–917.

Fitzmaurice, G. M. and Laird, N. M. (1993) A likelihood-based method for analysing longitudinal binary responses. *Biometrika*, **80**, 141–151.

Folstein, M. F., Folstein, S. E. and McHugh, P. R. (1975) Mini-mental state: a practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.*, **12**, 189–198.

Goldberg, D. (1972) *GHQ: the Selection of Psychiatric Illness by Questionnaire*. London: Oxford University Press.

Hall, C. B. (1995) Diagnosis for dependent data regression models. *PhD Thesis*. Department of Biostatistics, Johns Hopkins University, Baltimore.

Heagerty, P. J. and Zeger, S. L. (1996) Marginal regression models for clustered ordinal measurements. *J. Am. Statist. Ass.*, **91**, 1024–1036.

Hosmer, D. W. and Lemeshow, S. (1989) *Applied Logistic Regression*. New York: Wiley.

Jacquez, J. A., Mather, F. J. and Crawford, C. R. (1968) Linear regression with non-constant, unknown error variances: sampling experiments with least squares, weighted least squares and maximum likelihood estimators. *Biometrics*, **24**, 607–626.

Katz, S., Ford, A. B., Moskowitz, R. W., Jackson, B. A. and Jaffer, M. W. (1963) Studies of illness in the age. The index of ADL: a standardized measure of biological and psychosocial function. *J. Am. Med. Ass.*, **185**, 914–918.

Laird, N. M. and Ware, J. H. (1982) Random-effects models for longitudinal data. *Biometrics*, **38**, 963–974.

Landwehr, J. M., Pregibon, D. and Shoemaker, C. (1984) Graphical methods for assessing logistic regression models. *J. Am. Statist. Ass.*, **79**, 61–71.

Larsen, W. A. and McCleary, S. J. (1972) The use of partial residual plots in regression analysis. *Technometrics*, **14**, 781–790.

Lawton, M. P. and Brody, E. M. (1969) Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontologist*, **9**, 179–186.

Liang, K. Y. and Zeger, S. L. (1986) Longitudinal data and analysis using generalized linear models. *Biometrika*, **73**, 13–22.

———(1993) Regression analysis for correlated data. *A. Rev. Publ. Hlth*, **14**, 43–68.

Liang, K.-Y., Zeger, S. L. and Qaqish, B. (1992) Multivariate regression analyses for categorical data (with discussion). *J. R. Statist. Soc.* B, **54**, 3–40.

Lipsitz, S. R., Laird, N. M. and Harrington, D. P. (1991) Generalized estimating equations for correlated binary data: using the odds ratio as a measure of association. *Biometrika*, **78**, 153–160.

Little, R. J. A. and Rubin, D. B. (1987) *Statistical Analysis with Missing Data*. New York: Wiley.

Mangione, C. M., Phillips, R. S., Seddon, J. M., Lawrence, M. G., Cook, E. F., Dailey, R. and Goldman, L. (1992) Development of the ''Activities of Daily Vision'' scale: a measurement of visual functional status. *Med. Care*, **30**, 1111–1126.

McCullagh, P. (1980) Regression models for ordinal data (with discussion). *J. R. Statist. Soc.* B, **42**, 109–142.

McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models*, 2nd edn. London: Chapman and Hall.

Morrison, D. F. (1990) *Multivariate Statistical Methods*, 3rd edn. New York: McGraw-Hill.

O'Hara Hines, R. J. and Carter, E. M. (1993) Improved added variable and partial residual plots for the detection of influential observation in generalized linear models (with comments). *Appl. Statist.*, **42**, 3–20.

Pelli, D. G., Robson, J. G. and Wilkins, A. J. (1988) The design of a new letter chart for measuring contrast sensitivity. *Clin. Vis. Sci.*, **2**, 187–199.

Prentice, R. L. and Zhao, L. P. (1991) Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics*, **47**, 825–839.

Rubin, G. S., Bandeen-Roche, K., Orasadarao, P. and Fried, L. P. (1994) Visual impairment and disability in older adults. *Optom. Vis. Sci.*, **71**, 750–760.

Rubin, G. S., West, S. K., Munoz, B., Bandeen-Roche, K., Zeger, S. L., Schein, O. and Fried, L. P. (1997) A comprehensive assessment of visual impairment in an older american population: SEE study. *Invest. Ophth. Vis. Sci.*, **38**, 557–568.

Scheffé, H. (1959) *The Analysis of Variance*. New York: Wiley.

Steinberg, E. P., Tielsch, J. M., Schein, O. D., Javitt, J. C., Sharkey, P., Cassard, S. D., Legro, M. W., Dienerwest, M., Bass, E. B., Damiano, A. M., Steinwachs, D. M. and Sommer, A. (1994) The VF-14: an index of functional impairment in patients with cataract. *Arch. Ophth.*, **112**, 630–638.

Stewart, A. L. and Ware, J. E. (1992) *Measuring Functioning and Well-being: the Medical Outcomes Study Approach*, pp. 73–85. Durham: Duke University Press.

Stiratelli, R., Laird, N. and Ware, J. H. (1984) Random-effects models for serial observations with binary response. *Biometrics*, **40**, 961–971.

Ware, J. H. (1985) Linear models for the analysis of several measurements in longitudinal studies. *Am. Statistn*, **39**, 95–101.

West, S. K., Munoz, B., Rubin, G. S., Schein, O. D., Bandeen-Roche, K., Zeger, S. L., German, P. S. and Fried, L. P. (1997) Function and visual impairment in a population-based study of older adults: SEE project. *Invest. Ophth. Vis. Sci.*, **38**, 72–82.

Zellner, A. (1962) An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *J. Am. Statist. Ass.*, **57**, 348–368.

Zhao, L. P. and Prentice, R. L. (1990) Correlated binary regression using a quadratic exponential model. *Biometrika*, **77**, 642–648.