Biometrics 73, 214–219
March 2017

DOI: 10.1111/biom.12565

Ordinal Probability Effect Measures for Group Comparisons in Multinomial Cumulative Link Models

Alan Agresti[1,*] and Maria Kateri[2,]**

[1]Department of Statistics, University of Florida, Gainesville, Florida 32605, U.S.A.
[2]Institute of Statistics, RWTH Aachen University, D-52056 Aachen, Germany
[*]*email:* aa@stat.ufl.edu
[**]*email:* maria.kateri@rwth-aachen.de

Summary. We consider simple ordinal model-based probability effect measures for comparing distributions of two groups, adjusted for explanatory variables. An "ordinal superiority" measure summarizes the probability that an observation from one distribution falls above an independent observation from the other distribution, adjusted for explanatory variables in a model. The measure applies directly to normal linear models and to a normal latent variable model for ordinal response variables. It equals $\Phi(\beta/\sqrt{2})$ for the corresponding ordinal model that applies a probit link function to cumulative multinomial probabilities, for standard normal cdf Φ and effect β that is the coefficient of the group indicator variable. For the more general latent variable model for ordinal responses that corresponds to a linear model with other possible error distributions and corresponding link functions for cumulative multinomial probabilities, the ordinal superiority measure equals $\exp(\beta)/[1+\exp(\beta)]$ with the log–log link and equals approximately $\exp(\beta/\sqrt{2})/[1+\exp(\beta/\sqrt{2})]$ with the logit link, where β is the group effect. Another ordinal superiority measure generalizes the difference of proportions from binary to ordinal responses. We also present related measures directly for ordinal models for the observed response that need not assume corresponding latent response models. We present confidence intervals for the measures and illustrate with an example.

Key words: Cumulative logit model; Cumulative probit model; Mann–Whitney statistic; Ordinal multinomial models; Proportional odds; Stochastic ordering.

1. Introduction

This article considers simple ordinal effect summaries for model-based comparison of two groups on an ordinal categorical response variable, while adjusting for other explanatory variables. Unlike standard summaries using nonlinear measures such as probits and odds ratios that can be difficult for practitioners to interpret, the proposed measures are based merely on probabilities and their differences.

The summary measures generalize two "ordinal superiority" measures that compare two groups without supplementary explanatory variables. Let y_1 and y_2 denote independent random variables from groups denoted by A and B, for a quantitative or ordinal categorical scale. The measure

$$\Delta = P(y_1 > y_2) - P(y_2 > y_1). \qquad (1)$$

summarizes their relative size. For binary responses with outcomes (0, 1), this simplifies to the difference of proportions, $P(y_1 = 1) - P(y_2 = 1)$. If y_1 and y_2 are identically distributed, then $\Delta = 0.0$. For discrete response variables, such as ordinal categorical responses, a related measure that has null value equal to 0.50 rather than 0 is

$$\gamma = P(y_1 > y_2) + \frac{1}{2}P(y_1 = y_2) \qquad (2)$$

(Klotz, 1966). The correction factor adjusts for ties to generate a null value of 0.50. The measures are functionally related,

$$\gamma = (\Delta + 1)/2, \quad \Delta = 2\gamma - 1,$$

with γ and Δ having ranges [0, 1] and [−1, 1], respectively. They are most meaningful when the groups are stochastically ordered, such as when they differ by a location shift on some scale. For details for ordinal categorical response scales, see Agresti (2010, Chap. 2). The measures relate directly to the information used in the Mann–Whitney statistic. For example, issue 4 of Volume 25 of *Statistics in Medicine* in 2006, which is devoted to that statistic and its uses and extensions, contains several articles that use such measures.

The ordinal effect measures discussed in this article use such probabilities in the context of modeling ordinal response variables while adjusting for explanatory variables. Section 2 introduces the measures for normal linear models that contain an indicator term for the groups, because linear models serve as latent variable models for ordinal response data. Section 3 presents related measures for a standard model for an ordinal response variable that applies a link function such as the probit or logit to cumulative probabilities, utilizing its connection with the latent variable model for various error distributions. Section 4 presents an example, also showing how to use

R software to easily construct confidence intervals for the measures. Section 5 presents related ordinal effect measures for cumulative link models in terms of the observed response, instead of a latent response. Section 6 discusses the applicability of the measures and suggests extensions for other models.

2. Ordinal Superiority Measures for Normal Linear Models

We first consider normal linear models that have explanatory variables in addition to a binary group indicator variable. At explanatory variable values $\boldsymbol{x} = (x_1, \ldots, x_p)^T$, let y_1 denote the response variable for an observation in group A and let y_2 denote an independent response for an observation in group B. Using the model-based conditional distributions on y for the two groups at $\boldsymbol{x}$, let

$$\Delta = P(y_1 > y_2; \boldsymbol{x}) - P(y_2 > y_1; \boldsymbol{x}).$$

With no explanatory variables other than the group indicator, this simplifies to (1). An analog of the ordinal superiority measure (2) is

$$\gamma = (\Delta + 1)/2,$$

which is merely $P(y_1 > y_2; \boldsymbol{x})$ when the response is continuous. The measures are useful summaries when no substantive interaction occurs between the group variable and the explanatory variables.

Let z be a group indicator for an observation, where $z = 1$ for group A and $z = 0$ for group B. These ordinal measures have simple form for the ordinary normal linear model

$$y = \beta_0 + \beta z + \boldsymbol{x}^T \boldsymbol{\beta}_x + \epsilon,$$

with $\boldsymbol{\beta}_x = (\beta_1, \ldots, \beta_p)^T$ and $\epsilon \sim N(0, \sigma^2)$. For this model, the difference between the conditional means of y_1 and y_2 at $\boldsymbol{x}$ is β, and

$$\gamma = P(y_1 > y_2; \boldsymbol{x}) = P\left[\frac{(y_1 - y_2) - \beta}{\sqrt{2}\sigma} > \frac{-\beta}{\sqrt{2}\sigma}\right] = \Phi\left(\frac{\beta}{\sqrt{2}\sigma}\right).$$

This formula applies regardless of the values $\boldsymbol{x}$ of the explanatory variables. Likewise, $\Delta = 2\Phi(\beta/\sqrt{2}\sigma) - 1$. Differences between the normal conditional standardized means for the two groups taking values β/σ equal to 0, 0.5, 1, 2, 3, correspond to γ equal to 0.50, 0.64, 0.76, 0.92, 0.98, respectively. Analogous measures apply when interaction occurs between the group indicator and an explanatory variable, or when the variance is allowed to be nonconstant, but then the values of the measures depend on the value of that explanatory variable. The standardized difference β/σ has seen longtime use in the literature for comparing two groups (e.g., Lehmann, 1975, p. 71). The corresponding ordinal superiority measures have also been used in a general regression context (e.g., Brumback et al., 2006, Thas et al., 2012).

In practice, with least squares estimate $\hat{\beta}$ in the linear model and residual standard deviation s, we can estimate the ordinal group comparisons by $\hat{\gamma} = \Phi(\hat{\beta}/\sqrt{2}s)$ and $\hat{\Delta} = 2\Phi(\hat{\beta}/\sqrt{2}s) - 1$. A confidence interval (L, U) for the standardized difference β/σ in the normal linear model yields a corresponding confidence interval $(\Phi(L/\sqrt{2}),\ \Phi(U/\sqrt{2}))$ for γ, which then also yields one for Δ. For the model matrix $\boldsymbol{X}$ for the linear model, let v denote the element in the row and column of $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ corresponding to the effect parameter β for comparing the two groups. For testing H_0: $\beta = 0$ using the usual t statistic, $t = \hat{\beta}/s\sqrt{v}$, consider the noncentrality parameter

$$\lambda = \frac{\beta}{\sigma\sqrt{v}}.$$

Let $(\hat{\lambda}_L, \hat{\lambda}_U)$ denote the standard confidence interval for λ for this test (Lehmann, 1986, p. 352). Then, since $\lambda = (\beta/\sqrt{2}\sigma)(\sqrt{2/v})$, it follows that the confidence interval (L, U) for $\beta/\sqrt{2}\sigma$ is $\sqrt{v/2}(\hat{\lambda}_L, \hat{\lambda}_U)$. Applying Φ to these endpoints yields the confidence interval for γ. Hayter (2012) presented more general confidence intervals, and Tian (2008) presented confidence intervals for group comparisons when the groups have different variances.

3. Ordinal Superiority Measures for Ordinal Latent Variable Models

When y is a c-category ordinal response variable, the most popular models are special cases of the *cumulative link model*

$$\text{link}[P(y \le j)] = \alpha_j - \beta z - \boldsymbol{x}^T \boldsymbol{\beta}_x\ , \quad j = 1, \ldots, c-1, \qquad (3)$$

for link functions such as the logit, probit, or log–log and complementary log–log (McCullagh, 1980). It is often sensible to regard an ordinal categorical variable as necessarily crude measurement of a continuous latent variable y^* that, if we could observe it, would be the response variable in an ordinary linear model. The cumulative link model is implied by a model in which a latent response has conditional distribution with cdf given by the inverse of the link function and with mean $\beta z + \boldsymbol{x}^T \boldsymbol{\beta}_x$ (Anderson and Philips, 1981).

The normal latent variable model with $y^* \sim N(\beta z + \boldsymbol{x}^T \boldsymbol{\beta}_x, 1)$ implies the cumulative probit model

$$\Phi^{-1}[P(y \le j)] = \alpha_j - \beta z - \boldsymbol{x}^T \boldsymbol{\beta}_x,$$

with $\{\alpha_j\}$ being cutpoints on the underlying scale and Φ being the standard normal cdf. The ordinal superiority measures apply directly to this latent variable model. Let y_1^* and y_2^* denote independent underlying latent variables at $\boldsymbol{x}$ when $z = 1$ and when $z = 0$, respectively. For this model,

$$\gamma = P(y_1^* > y_2^*; \boldsymbol{x}) = P\left[\frac{(y_1^* - y_2^*) - \beta}{\sqrt{2}} > \frac{-\beta}{\sqrt{2}}\right] = \Phi\left(\frac{\beta}{\sqrt{2}}\right),$$

regardless of $\boldsymbol{x}$ values, and $\Delta = 2\Phi(\beta/\sqrt{2}) - 1$.

The logit link and corresponding cumulative logit model relate to underlying logistic distributions, for which such a simple expression does not occur. However, because of the very close similarity of logit and probit model fits, estimates of the corresponding measures for that logistic latent variable model are very similar to estimates for the normal latent variable model. For a cumulative logit model with proportional odds structure and maximum likelihood estimate $\hat{\beta}$ of the group effect, we can use numerical integration or simulate pairs of observations from the relevant logistic distributions to closely approximate the maximum likelihood estimate of the probability for the difference of latent logistic random variables. In practice, though, it is adequate to approximate the distribution of $y_1^* - y_2^*$ by a logistic distribution with parameter β and scale parameter $\sqrt{2}$, for which

$$\gamma \approx \frac{\exp(\beta/\sqrt{2})}{[1+\exp(\beta/\sqrt{2})]},$$

or to fit the corresponding cumulative probit model and use the closed-form results for it.

For ordinal responses, log–log and complementary log–log links are appropriate when we expect underlying latent variables to have extreme-value distributions. If in the latent variable model, the errors are independent extreme-value random variables (i.e., the standard Gumbel cdf $F(\epsilon) = \exp[-\exp(-\epsilon)]$), then their difference has the standard logistic distribution (McFadden, 1974). For a model with log–log link and coefficient β for the group indicator, it follows that

$$\gamma = P(y_1^* > y_2^*; \boldsymbol{x}) = \frac{\exp(\beta)}{[1+\exp(\beta)]},$$

when the scale parameter of the underlying extreme-value distributions is 1.

For γ and Δ for the latent variable model with an ordinal response variable, simple confidence intervals result directly from ordinary confidence intervals for β for the corresponding ordinal cumulative link model. For example, if $[\hat{\beta}_L, \hat{\beta}_U]$ is a profile-likelihood or Wald confidence interval for β in the cumulative probit model based on a multinomial likelihood, the corresponding confidence interval for γ is $[\Phi(\hat{\beta}_L/\sqrt{2}), \Phi(\hat{\beta}_U/\sqrt{2})]$.

4. Example for Cumulative Link Models

We illustrate the ordinal superiority measures with an example from Agresti (2015, Section 6.3.3) on a study of mental health. It relates a four-category response variable measuring mental impairment (1 = well, 2 = mild symptom formation, 3 = moderate symptom formation, 4 = impaired) to a binary indicator of socioeconomic status (SES: 1 = high, 0 = low) and a quantitative life-events (LE) index taking values on the nonnegative integers between 0 and 9 with mean 4.3 and standard deviation 2.7. The $n = 40$ observations are available at www.stat.ufl.edu/ aa/glm/data.

For the cumulative probit model corresponding to a normal latent variable model, the maximum likelihood fit is

$$\Phi^{-1}[\hat{P}(y \le j)] = \hat{\alpha}_j + 0.68336(\text{SES}) - 0.19535(\text{LE}).$$

To compare the two levels of SES using $\hat{\beta}_1 = -0.68336$, we can use $\hat{\gamma} = \Phi(\hat{\beta}_1/\sqrt{2}) = 0.314$ and $\hat{\Delta} = -0.371$. The ordinal superiority measure $\hat{\gamma}$ has the interpretation that at any particular value for life events, there is about a 1/3 chance of lower mental impairment at low SES than at high SES. The 95% profile likelihood confidence interval for β_1 yields confidence intervals (0.161, 0.507) for γ and (−0.678, 0.015) for Δ. Table 1 shows how simple it is to use software such as `R` to obtain a confidence interval for γ for the SES effect. Here, we fitted the cumulative probit model using the *cml* function of the `R`–package `ordinal` (Christensen, 2011).

Similarly, we can use these measures to compare two levels of the life events measure. For the highest and lowest levels (0 and 9), $\hat{\gamma} = \Phi(9\hat{\beta}_2/\sqrt{2}) = 0.893$, with 95% profile likelihood confidence interval (0.653, 0.983), suggesting a very strong effect.

Table 1

R code and output (edited) for finding confidence interval for ordinal superiority measure γ for SES effect in cumulative probit model with mental impairment data

```
> Mental <- read.table("http://www.stat.ufl.edu/~aa/glm/data/Mental.dat",header=T)
> Mental
1  impair ses life
2       1   1    1
3       1   1    9
...
40      4   0    9
> attach(Mental)
> library(ordinal) # library(ordinal) requires response to be a factor
> impair.f <- factor(impair)
> probit.m <- clm(impair.f ~ ses + life, link="probit")
> summary(probit.m) # we don't show cutpoint parameter estimates
     Estimate Std. Error z value Pr(>|z|)
ses  -0.68336    0.36411  -1.877  0.06055 .
life  0.19535    0.06887   2.837  0.00456
> Like.CI.b1 <-  confint(probit.m)[1,]  # profile likelihood CI for beta1
> Like.CI.gamma <- pnorm(Like.CI.b1/sqrt(2)); Like.CI.gamma
    2.5 %    97.5 %
0.1608020 0.5074911
```

In practice, some methodologists use ordinary normal linear models for ordinal response data, feeling they are easier to interpret than cumulative link models and believing that the response is merely crude measurement of something inherently continuous. For comparison, we estimate the ordinal superiority measure γ with this crude modeling approach, showing results in Table 2 using `R` software. For the SES effect, $\hat{\gamma} = \Phi(\hat{\beta}_1/\sqrt{2}s) = \Phi[-0.64501/\sqrt{2}(1.02696)] = 0.328$. The confidence interval for the noncentrality parameter λ is available in R software with the *conf.limits.nct* function in the `MBESS` package (Kelley, 2007). So, it is simple to obtain a confidence interval for γ, using $v = (se/s)^2$, where *se* is the reported standard error for the estimated group effect. Table 2 shows that the 95% confidence interval for γ is (0.18, 0.51), quite similar to (0.16, 0.51) obtained with the truly ordinal treatment of the data through the cumulative probit model.

Next, we consider two alternative link functions for the cumulative link model. The cumulative logit model has $\hat{\beta} = -1.1112$ for the SES effect, for which $\hat{\gamma} = 0.317$. The value $\hat{\gamma} \approx \exp(\hat{\beta}/\sqrt{2})/[1 + \exp(\hat{\beta}/\sqrt{2})] = 0.313$ which has approximate 95% profile likelihood confidence interval (0.160, 0.511), nearly identical to what we obtain with the cumulative probit model. Using the log–log link for an underlying extreme-value distribution, which is plausible for mental impairment and quite different from a logit or probit link, we obtain $\hat{\beta} = -0.8746$ and $\hat{\gamma} = \exp(\hat{\beta})/[1 + \exp(\hat{\beta})] = 0.294$ with 95% profile likelihood confidence interval (0.152, 0.487). Comparison of log-likelihood or AIC values does not suggest a clear preference among the probit, logit, and log–log links, partly reflecting the modest sample size.

In yet another approach for these data, Thas et al. (2012) fitted a semiparametric model for logit(γ), estimating model parameters using a set of estimating equations and estimating the covariance matrix of the estimators with a sandwich covariance matrix. At fixed life events, they obtained $\hat{\gamma} = 0.32$ and a 95% confidence interval for γ of (0.20, 0.48). This is slightly narrower than obtained with parametric models but very similar to one we obtain below without assuming a latent variable model, and with similar substantive impact. We consider this mental impairment example further in the next section.

5. Measures for Ordinal Models without Assuming Latent Structure

For any model for ordinal categorical responses, analogs of the ordinal superiority measures apply directly to the model, without reference to any latent variable model. For a c-category ordinal response variable y and a specific value of the explanatory vector $\boldsymbol{x}$, say $\boldsymbol{x}_0$, let

$$\pi_{\ell j}(\boldsymbol{x}_0) = P(y = j; z = 2 - \ell, \boldsymbol{x}_0), \quad \ell = 1, 2, \quad j = 1, \ldots, c.$$

We define

$$\Delta(\boldsymbol{x}_0) = \sum_{j>k} \pi_{1j}(\boldsymbol{x}_0)\pi_{2k}(\boldsymbol{x}_0) - \sum_{k>j} \pi_{1j}(\boldsymbol{x}_0)\pi_{2k}(\boldsymbol{x}_0), \qquad (4)$$

and

$$\gamma(\boldsymbol{x}_0) = \sum_{j>k} \pi_{1j}(\boldsymbol{x}_0)\pi_{2k}(\boldsymbol{x}_0) + \frac{1}{2}\sum_{j} \pi_{1j}(\boldsymbol{x}_0)\pi_{2j}(\boldsymbol{x}_0). \qquad (5)$$

Corresponding sample values $\hat{\Delta}(\boldsymbol{x}_0)$ and $\hat{\gamma}(\boldsymbol{x}_0)$ replace the probabilities in (4) and (5) by the corresponding fitted values $\{\hat{\pi}_{1j}(\boldsymbol{x}_0)\}$ and $\{\hat{\pi}_{2j}(\boldsymbol{x}_0)\}$ for the model.

Unlike the measures for the latent variable models, these measures have values depending on $\boldsymbol{x}_0$. In practice, we could report them and their confidence intervals at a representative $\boldsymbol{x}_0$ value, such as the overall mean $\bar{\boldsymbol{x}}$. Or, if the sample $\boldsymbol{x}$ values are representative of the population of interest, a summary approach estimates the measures at the $\boldsymbol{x}$ value for each obser-

Table 2

R code and output (edited) for finding confidence interval for ordinal superiority measure γ for SES effect in normal linear model with mental impairment data

```
> summary(lm(impair ~ ses + life))
Coefficients:
               Estimate  Std. Error  t value   Pr(>|t|)
(Intercept)  1.91973788  0.33785712  5.68210 1.6924e-06
ses         -0.64500836  0.32915094 -1.95961  0.0576069 .
life         0.17778169  0.06060938  2.93324  0.0057285
---
Residual standard error: 1.02696 on 37 degrees of freedom

> library(MBESS)
> conf.limits.nct(-0.64501/0.32915, df=37, conf.level=0.95)
$Lower.Limit
[1] -3.956887716
$Upper.Limit
[1] 0.06278409435
> v <- (0.32915/1.02696)^2
> pnorm(sqrt(v/2)*(-3.956887)); pnorm(sqrt(v/2)*(0.062784))
[1] 0.1849219812
[1] 0.5056763573
```

vation, and then averages them. Let $\boldsymbol{x}_i$ denote the explanatory component vector for observation i, and let $\pi_{\ell ij} = \pi_{\ell j}(\boldsymbol{x}_i)$, for $\ell = 1, 2$, $i = 1, \ldots, n$, and $j = 1, \ldots, c$. Summary ordinal superiority measures are

$$\Delta^* = \frac{1}{n}\sum_i \Delta_i \quad \text{and} \quad \gamma^* = \frac{1}{n}\sum_i \gamma_i, \tag{6}$$

with components $\Delta_i = \Delta(\boldsymbol{x}_i)$ and $\gamma_i = \gamma(\boldsymbol{x}_i)$, given by (4) and (5), respectively. The expressions of Δ_i and γ_i in terms of the parameters of the cumulative link model (3) are given in web appendix A. We obtain the model-based estimates $\hat{\Delta}^*$ and $\hat{\gamma}^*$ by replacing the parameter values by the corresponding estimated values.

To construct confidence intervals for these measures, we can obtain large-sample standard errors using the delta method, based on an estimated covariance matrix of the ML model parameter estimates that are generated from the usual multinomial sampling scheme. From results for the simple case without explanatory variables, it is more sensible to apply the delta method to a transform such as the logit of the measure (Ryu and Agresti, 2008) rather than to the measure itself. Web appendix A contains the technical details. An `R` function for constructing estimates and confidence intervals for γ^* and Δ^*, based on the cumulative logit or probit model, is available in web appendix B.

We illustrate for the mental impairment data that Section 4 used to illustrate the measures for the ordinal latent variable models. For comparing the two SES levels with cumulative probit and cumulative logit models, Table 3 shows $\gamma(x)$ and $\Delta(x)$ at the life events index values $x = 0, \ldots, 9$, and at the sample mean value $\bar{x} = 4.3$. Although the estimates vary according to the life events value, they are quite stable. As we would expect, because of the similarity of logit and probit models, summary results are similar for the two cumulative links.

Table 3

Estimates of the ordinal superiority measures comparing the two SES levels for the mental impairment data at the different levels of the life-events index and its sample mean, based on the cumulative probit and cumulative logit models

	Cumulative			
	Probit	Logit	Probit	Logit
Life events	$\hat{\gamma}$		$\hat{\Delta}$	
0	0.355	0.357	−0.291	−0.286
1	0.345	0.348	−0.310	−0.305
2	0.338	0.341	−0.325	−0.318
3	0.333	0.337	−0.334	−0.326
4	0.330	0.335	−0.340	−0.330
5	0.329	0.334	−0.342	−0.333
6	0.330	0.334	−0.339	−0.332
7	0.334	0.336	−0.333	−0.327
8	0.339	0.341	−0.321	−0.317
9	0.348	0.350	−0.305	−0.301
$\bar{x} = 4.3$	0.330	0.334	−0.341	−0.331

For the summary measures averaged over the 40 observations, we obtain $\hat{\gamma}^* = 0.337$ and $\hat{\Delta}^* = -0.325$ for the probit model, and we obtain $\hat{\gamma}^* = 0.341$ and $\hat{\Delta}^* = -0.319$ for the logit model. Table 4 shows 95% confidence intervals for the population values, using the observed information matrix. All these analyses indicate a range from essentially no effect to a relatively large one in the direction of poorer mental health at the lower SES level.

6. Discussion and Extensions

The measures introduced here supplement measures previously proposed to summarize effects in models for ordinal categorical responses, such as Ryu and Agresti (2008) and Thas et al. (2012). For other ordinal effect measures, see Cheng (2009), Lu et al. (2014), Lu et al. (2015), and Volfovsky et al. (2015).

An advantage of the ordinal superiority measures is simplicity of interpretation for ordinal categorical models in which researchers often find probits and odds ratios difficult to interpret. For models with nonlinear link functions, such as cumulative link models, the natural model-based effect measures are not easy to understand. For the typical medical researcher or practitioner, for instance, reading that at any values of explanatory variables the estimated probability that a response to drug ($z = 1$) is better than a response to placebo ($z = 0$) is $\hat{\gamma} = 0.66$ would have greater meaning than reading that (i) an estimated cumulative odds for drug is $\exp(\hat{\beta}) = 2.7$ times the estimated cumulative odds for placebo (i.e., from (3) with the logit link), or (ii) estimated cumulative probits differ by $\hat{\beta} = 0.5$ or an underlying mean for drug is $\hat{\beta} = 0.5$ standard deviations better than for placebo (i.e., from (3) with the probit link), or (iii) the estimated probability that the response for drug is worse than a particular outcome category is the power $\exp(\hat{\beta}) = 1.7$ of the estimated probability that the response for placebo is worse than that category (i.e., from (3) with the complementary log–log link).

The ordinal superiority measures extend directly to summary comparisons of multiple groups, based on more general models that have multiple indicator variables for the groups. For example, suppose a cumulative probit model contains terms $\beta^{(a)}z_a + \beta^{(b)}z_b$ in the linear predictor for groups a and b, where $z_j = 1$ for observations from group j and $z_j = 0$ otherwise. Then, an analog of γ for comparing those groups is $\Phi[(\beta^{(a)} - \beta^{(b)})/\sqrt{2}]$. Inference can use Bonferroni adjustments. With a large number g of groups, it may be useful to model the $g(g-1)/2$ comparison measures in terms of fewer parameters, such as is done with the Bradley–Terry model and is discussed in a simpler context by Bergsma et al. (2009, p. 11).

The proposed measures in Section 5 that are not connected with a linear latent variable model apply directly to other ordinal models, such as continuation-ratio logit models and adjacent-category logit models that have proportional odds structure (Agresti, 2010, Chapter 4). When the explanatory variables are solely categorical, the data form a contingency table, and (3) for the logit link is the response model analog of association models for cumulative odds ratios, while other

Table 4

95% confidence intervals for the ordinal superiority measures comparing the two SES levels for the mental impairment data at the sample mean of the life-events index and summarized over life-events values, based on the cumulative probit and cumulative logit models

	Cumulative			
	Probit	Logit	Probit	Logit
Life events	γ		Δ	
$\bar{x} = 4.3$	(0.19, 0.51)	(0.20, 0.51)	(−0.63, 0.03)	(−0.61, 0.02)
Summary	(0.21, 0.49)	(0.21, 0.50)	(−0.57, −0.02)	(−0.57, −0.01)

ordinal response models correspond to association models for alternative types of ordinal odds ratios (see Sections 8.3.2–8.3.4 of Kateri, 2014). Some of these models, such as those expressed in terms of local odds ratios, have approximate connections with underlying normal models. The measures extend also to more general ordinal-response models than those having linear predictors, such as generalized additive models for ordinal responses (e.g., Yee and Wild, 1996), although obtaining confidence intervals is then more challenging.

7. Supplementary Materials

Web Appendices A and B, referenced in Section 5, are available with this article at the *Biometrics* website on Wiley Online Library. Web appendix A contains the technical details for deriving the large-sample confidence intervals for γ^* and Δ^*, while web appendix B provides the R–function for computing $\hat{\gamma}^*$ and $\hat{\Delta}^*$, along with the associated confidence intervals.

Acknowledgements

The authors appreciate helpful comments about an earlier draft from Wicher Bergsma, Leonardo Grilli, Carla Rampichini, and Euijung Ryu.

References

Agresti, A. (2010). *Analysis of Ordinal Categorical Data*, 2nd ed. Hoboken, NJ: Wiley.

Agresti, A. (2015). *Foundations of Linear and Generalized Linear Models*. Hoboken, NJ: Wiley.

Anderson, J. A. and Philips, P. R. (1981). Regression, discrimination, and measurement models for ordered categorical variables. *Applied Statistics* **30**, 22–31.

Bergsma, W., Croon, M. A., and Hagenaars, J. A. (2009). *Marginal Models for Dependent, Clustered, and Longitudinal Categorical Data*. New York, NY: Springer.

Brumback, L. C., Pepe, M. S., and Alonzo, T. A. (2006). Using the ROC curve for gauging treatment effect in clinical trials. *Statistics in Medicine* **25**, 575–590.

Cheng, J. (2009). Estimation and inference for the causal effect of receiving treatment on a multinomial outcome. *Biometrics* **65**, 96–103.

Christensen, R. H. B. (2011). Analysis of ordinal data with cumulative link models estimation with the `ordinal` package. *R–package version 2011.09-13*.

Hayter, A. J. (2012). Win-probabilities for regression models. *Statistical Methodology* **9**, 520–527.

Kateri, M. (2014). *Contingency Table Analysis: Methods and Implementation Using R*. New York: Birkäuser/Springer.

Kelley, K. (2007). Confidence intervals for standardized effect sizes: theory, application, and implementation. *Journal of Statistical Software* **20**, 1–24.

Klotz J. H. (1966). The Wilcoxon, ties, and the computer. *Journal of the American Statistical Association* **61**, 772–787.

Lehmann, E. L. (1975). *Nonparametrics: Statistical Methods Based on Ranks*. San Francisco, CA: Holden-Day.

Lehmann, E. L. (1986). *Testing Statistical Hypotheses*, 2nd edition. New York, NY: Springer-Verlag.

Lu, J., Ding, P., and Dasgupta, T. (2015). Sharp bounds of causal effects on ordinal outcomes. http://arxiv.org/abs/1507.01542v1

Lu, T.-Y., Poon, W.-Y., and Cheung, S. H. (2014). A unified framework for the comparison of treatments with ordinal responses. *Pscyhometrika* **79**, 605–620.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society, Series B* **42**, 109–142.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics*, P. Zarembka (ed), 105–142. New York: Academic Press.

Ryu, E. and Agresti, A. (2008). Modeling and inference for an ordinal effect size measure. *Statistics in Medicine* **27**, 1703–1717.

Thas, O., De Neve, J., Clement, L., and Ottoy, J.-P. (2012). Probabilistic index models. *Journal of the Royal Statistical Society, Series B* **74**, 623–671.

Tian, L. (2008). Confidence intervals for $P(Y_1 > Y_2)$ with normal outcomes in linear models. *Statistics in Medicine* **27**, 4221–4237.

Volfovsky, A., Airoldi, E. M., and Rubin, D. B. (2015). Causal inference for ordinal outcomes. http://arxiv:1501.01234v1

Yee, T. W. and Wild, C. J. (1996). Vector generalized additive models. *Journal of the Royal Statistical Society, Series B* **58**, 481–493.

Received February 2016. *Revised June* 2016.
Accepted June 2016.