Thai **J**ournal of **M**athematics : (2016) 63–70
Special Issue on Applied Mathematics : Bayesian Econometrics

http://thaijmath.in.cmu.ac.th
ISSN 1686-0209

Causal Effect for Ordinal Outcomes from Observational Data: Bayesian Approach

Jirakom Sirisrisakulchai[1] **and Songsak Sriboonchitta**

Faculty of Economics, Chiang Mai University
e-mail : sirisrisakulchai@hotmail.com (J. Sirisrisakulchai)
songsakecon@gmail.com (S. Sriboonchitta)

Abstract : Ordinal outcomes are often observed in the social and economic sciences. It is frequently that the scale or magnitude of the outcomes is not available. The common average treatment effect is not well-defined for causal inference. We define a useful causal estimands for ordinal outcomes in this research. To consistently estimate the causal estimands, the data has to satisfy the ignorable treatment assignment assumption. This condition ensures that the outcome of interest is independent of the treatment assignment mechanism. We discuss and propose the models for correcting self-selection bias from this type of observed data using copula approach. Copula can capture the dependence between treatment assignment and outcomes of interest. Bayesian estimation procedures play an important role in causal analysis [1]. Thus, Bayesian estimation procedure is applied to help estimating the complex model structures. Finally, we discuss the framework for estimate causal effect of ordinal potential outcomes and apply this framework to the healthcare survey data from [2] as a case study.

Keywords : Bayesian causal inference; ordinal outcomes; observational data.
2010 Mathematics Subject Classification : 62P20; 62P30.

1 Introduction

One of the main goals of scientific research is mostly to investigate the causal relationship between treatments and outcomes of interest. In general, it is very complex and challenging to define cause. However, for empirical research, the definition of causal effect of treatment is straightforward and useful. [3] defined causal effect through potential outcomes. Causal estimands are simply to compare the potential outcomes that would have been observed under different exposures of units to treatment [4]. [3] discussed the potential outcomes framework for causal inference on the binary outcome. This work

This research was supported in part by the Thailand Research Fund grant MRG-5980209.
[1]Corresponding author.

provided a formal framework for the estimation of causal effects for many types of data. However, there is no explicit discussion in the recent literatures for ordinal outcomes.

Ordinal outcomes are frequently observed in the social and economic sciences. Sometimes, the scale or magnitude of the ordinal outcomes is not available. For example, level of education consists of high school, bachelor's degree, master's degree, and doctoral degree. This variable can be viewed as an ordinal variable but with no scale or magnitude between each ordering. If one would like to make a causal statement about how the government program to promote higher education affects the participants' level of education, one has to specify a causal estimand for this problem. The common estimand, such as the average treatment effect, is not well-defined for non-numeric and ordinal data in this case. The average of more than two ordinal values is difficult to interpret.

If we can observe both the potential outcomes under treatment and control or non-treatment for each individual or unit, then the causal effect is the difference between these potential outcomes. However, we can only observe one of the potential outcomes for each individual or unit. To estimate the causal effect, we need data for inference. To consistently estimate the treatment effect, we have to satisfy the ignorable treatment assignment assumption [5]. This condition ensures that the outcome of interest is independent of the treatment assignment mechanism [6]. This assumption can be satisfied when we assign each individual randomly to the treatment and control groups. In other words, we can only consistently estimate the treatment effect if there is no selection bias in the treatment assignment.

In some observational studies, the survey participants sometimes self-selected themselves into the treatment condition, thus it is hard to believe that the independence assumption holds in this situation. In this observed data, the treatment assignment is not ignorable, thus using treatment condition as the dummy variable in statistical models will lead to the bias estimation of treatment effect. Factors affecting self-selection might cause the dummy variable of treatment condition to be correlated with random errors in the outcome model, which leads to biased parameters estimation. Bayesian estimation procedures play an important role in causal analysis [1]. The objective of this paper is to propose a model that can treat this self-selection bias by allowing dependence between random errors of selection and outcome models using copula approach and applied Bayesian estimation procedure to estimate the parameters of the models.

From the proposed model, we can impute the counterfactual potential outcomes for each individual. This will give us the full matrix of joint distribution of potential outcomes. Based on the joint distribution matrix, we can calculate the causal effect defined as a function of this joint distribution. This procedure for causal inference will be applied to the healthcare survey data from [2] as a case study.

2 Model Formulation

In this section, we develop a parametric model to investigate the effect of a binary treatment variable on an ordinal outcome of interest. Let D_i indicate a binary treatment variable (or assignment) for unit i, where $D_i = 1$ implies the receipt of treatment and $D_i = 0$ implies the non-receipt or control. The ordinal outcome of interest is denoted $Y_i \in \{1, 2, ..., J\}$ for unit i. We use superscript to separate the potential outcomes received by the unit in the treatment and the control states, where $Y_i^{(1)}$ is the outcome of the unit under treatment and $Y_i^{(0)}$ is the outcome of the unit without treatment. We only observe one outcome Y_i for each unit and never observe both, and thus

$$Y_i = D_i Y_i^{(1)} + (1 - D_i) Y_i^{(0)}. \tag{2.1}$$

We model the observed treatment assignment D_i and the potential ordinal outcomes $Y_i^{(1)}$ and $Y_i^{(0)}$ by a latent variable framework as follows:

$$D_i^* = W_i \beta^{(D)} + u_i, \tag{2.2a}$$
$$Z_i^{(1)} = X_i \beta^{(1)} + \epsilon_i^{(1)}, \tag{2.2b}$$
$$Z_i^{(0)} = X_i \beta^{(0)} + \epsilon_i^{(0)}, \tag{2.2c}$$

where W_i and X_i are the vectors of covariate; $\beta^{(D)}$, $\beta^{(1)}$, and $\beta^{(0)}$ are the vectors of parameter; and u_i, $\epsilon_i^{(1)}$, and $\epsilon_i^{(0)}$ are the error terms. We assume an exclusion restriction for W_i and X_i, where some covariates in W_i are not contained in X_i [7].

The binary treatment variable D_i and the latent variable D_i^* are related as follows:

$$D_i = \begin{cases} 1 & \text{if } D_i^* > 0 \\ 0 & \text{if } D_i^* \leq 0. \end{cases}$$

Similarly, the potential outcomes $Y_i^{(1)}$ and $Y_i^{(0)}$ are related to the latent variables $Z_i^{(1)}$ and $Z_i^{(0)}$ as follows:

$$Y_i^{(k)} = j \quad \text{iff} \quad \tau_j^{(k)} < Z_i^{(k)} \leq \tau_{j+1}^{(k)}, k = 0, 1, j = 1, 2, ..., J, \tag{2.3}$$

where $\tau_j^{(k)}, k = 0, 1, j = 1, 2, ..., J$ are the threshold values, forming a partition of the real line, that will enable us to map the continuous latent values into discrete values of ordinal outcomes in both states. For the model identification purpose, we set $\tau_1^{(1)} = \tau_1^{(0)} = -\infty$, $\tau_2^{(1)} = \tau_2^{(0)} = 0$, and $\tau_{J+1}^{(1)} = \tau_{J+1}^{(0)} = \infty$.

Finally, the joint distribution of the error terms is modelled by copula function. A bivariate copula function with parameter θ, $C(v_1, v_2|\theta)$ is a distribution function over the interval $[0, 1] \times [0, 1]$ with uniform marginal distribution function [8]. For a bivariate joint distribution H with marginal distributions F_1 and F_2 , the copula $C : [0, 1]^2 \to [0, 1]$, which combines these two marginal distributions, can be expressed as follows:

$$H(x_1, x_2) = C\left(F_1(x_1), F_2(x_2)|\theta\right), (x_1, x_2) \in \mathbf{R^2}. \tag{2.4}$$

Given the assumed conditional independence across observations, the likelihood function for these models can be written as:

$$p(Y, D|\Gamma) \equiv L(\Gamma; Y, D) = \left(\prod_{i:D_i=1} \Pr(Y_i^{(1)} = j, D_i = 1|\Gamma) \right) \left(\prod_{i:D_i=0} \Pr(Y_i^{(0)} = j, D_i = 0|\Gamma) \right), \tag{2.5}$$

where $\Gamma = \{\beta^{(D)}, \beta^{(1)}, \beta^{(0)}, \tau^{(0)}, \tau^{(1)}, \theta^{(1)}, \theta^{(0)}\}$ is the vector of parameters to be estimated. Bivariate copula can be applied to calculate this likelihood as follows:

$$\begin{aligned} \Pr(Y_i^{(0)} = j, D_i = 0|\Gamma) &= \Pr(\tau_j^{(0)} - X_i\beta^{(0)} < \epsilon_i^{(0)} \leq \tau_{j+1}^{(0)} - X_i\beta^{(0)}, u_i \leq W_i\beta^{(D)}|W_i, X_i, \Gamma) \\ &= Pr(\epsilon_i^{(0)} \leq \tau_{j+1}^{(0)} - X_i\beta^{(0)}, u_i \leq W_i\beta^{(D)}|W_i, X_i, \Gamma) \\ &\quad - Pr(\epsilon_i^{(0)} \leq \tau_j^{(0)} - X_i\beta^{(0)}, u_i \leq W_i\beta^{(D)}|W_i, X_i, \Gamma), \end{aligned}$$

$$\begin{aligned}\Pr(Y_i^{(1)}=j, D_i=1|\Gamma) &= \Pr(\tau_j^{(1)} - X_i\beta^{(1)} < \epsilon_i^{(1)} \le \tau_{j+1}^{(1)} - X_i\beta^{(1)}, u_i > W_i\beta^{(D)}|W_i, X_i, \Gamma)\\ &= \Pr(\epsilon_i^{(1)} \le \tau_{j+1}^{(1)} - X_i\beta^{(1)}|W_i, X_i, \Gamma) - \Pr(\epsilon_i^{(1)} \le \tau_j^{(1)} - X_i\beta^{(1)}|W_i, X_i, \Gamma)\\ &\quad - Pr(\epsilon_i^{(1)} \le \tau_{j+1}^{(1)} - X_i\beta^{(1)}, u_i \le W_i\beta^{(D)}|W_i, X_i, \Gamma)\\ &\quad + Pr(\epsilon_i^{(1)} \le \tau_j^{(1)} - X_i\beta^{(1)}, u_i \le W_i\beta^{(D)}|W_i, X_i, \Gamma).\end{aligned}$$

For any given copula, the two required joint distributions, $\Pr(Y_i^{(1)} = j, D_i = 1|\Gamma)$ and $\Pr(Y_i^{(0)} = j, D_i = 0|\Gamma)$ are fully determined. Therefore,

$$\begin{aligned}\Pr(Y_i^{(0)}=j, D_i=0|\Gamma) &= C_0(F_1^{(0)}(\tau_{j+1}^{(0)} - X_i\beta^{(0)}), F_2(-W_i\beta^{(D)})|\theta^{(0)})\\ &\quad - C_0(F_1^{(0)}(\tau_j^{(0)} - X_i\beta^{(0)}), F_2(-W_i\beta^{(D)})|\theta^{(0)}),\end{aligned}$$

and

$$\begin{aligned}\Pr(Y_i^{(1)}=j, D_i=1|\Gamma) &= C_1(F_1^{(1)}(\tau_{j+1}^{(1)} - X_i\beta^{(1)}), 1|\theta^{(1)}) - C_1(F_1^{(1)}(\tau_j^{(1)} - X_i\beta^{(1)}), 1|\theta^{(1)})\\ &\quad - C_1(F_1^{(1)}(\tau_{j+1}^{(1)} - X_i\beta^{(1)}), F_2(-W_i\beta^{(D)})|\theta^{(1)})\\ &\quad + C_1(F_1^{(1)}(\tau_j^{(1)} - X_i\beta^{(1)}), F_2(-W_i\beta^{(D)})|\theta^{(1)}),\end{aligned}$$

where $C_0(v_1, v_2)$ and $C_1(v_1, v_2)$ are copula functions and $F_1^{(0)}$, $F_1^{(1)}$ and F_2 are marginal distribution functions.

3 Bayesian Estimation

To start a Bayesian estimation, we first specify the likelihood function for our ordinal outcome models discussed in the previous section. The likelihood function can be fully specified by assuming parametric models for the marginal distributions, $F_1^{(0)}$, $F_1^{(1)}$, F_2, and copula functions, C_0, C_1, for bivariate distribution models. We then add a prior density, $p(\Gamma)$, where Γ is the vector of parameters previously defined in Section 2. This prior represents a subjective belief about values of the parameters. If the researcher does not have a sufficient information on the parameters, this prior can be selected to be vague so that information contained in the data will dominate the vague information from the prior. By combining the prior density $p(\Gamma)$ with the likelihood $p(Y, D|\Gamma)$, we can obtain the joint posterior density $p(\Gamma|Y, D)$ from Bayes theorem. The joint posterior completely gives the full information for point and interval estimations of parameters including other quantities of interest, for instance, quantiles.

In practice, the direct evaluation of this posterior can be difficult involving a high-dimensional integration problem. The high-dimensional integration derived from our model formulation has no analytical solutions. To extract information from the posterior, we utilized a recent advances technology in simulation methods, namely, Gibbs sampler and Metropolis-Hastings algorithm. These algorithms make us tractable to generate a sequence of draws that converge to the posterior $p(\Gamma|Y, D)$. After achieving convergence, the set of simulated parameter values will be used to calculate the quantities of interest, for instance, posterior means. Gibbs sampler is performed by producing a Markov chain whose limiting distribution is $p(\Gamma|Y, D)$. This can be done by iteratively sampling from the complete posterior conditionals of the model. The Metropolis Hastings algorithm is a generalization of the Gibbs sampler. It can be viewed as a multivariate accept-reject algorithm. The readers are referred to [9] and [10] for a detailed reviews of these simulation methods.

In this paper, we perform the algorithms above with the use of data augmentation [11]. We first expand the posterior distribution by augmenting with the latent data D_i^*, $Z_i^{(1)}$, and $Z_i^{(0)}$. The use of data augmentations will simplify the required posterior calculation when used in conjunction with the Gibbs sampler [12]. For our model, the augmented posterior can be written as:

$$p(\Gamma, D^*, Z_i^{(1)}, Z^{(0)}|Y, D) = p(Y, D|\Gamma, D^*, Z_i^{(1)}, Z^{(0)}) \times p(D^*, Z_i^{(1)}, Z^{(0)}|\Gamma) \times p(\Gamma). \quad (3.1)$$

The middle term $p(D^*, Z_i^{(1)}, Z^{(0)}|\Gamma)$ is just the bivariate distributions of $(D^*, Z^{(1)})$ and $(D^*, Z^{(0)})$, when we assume the independence between $Z^{(1)}$ and $Z^{(0)}$. For the first term $p(Y, D|\Gamma, D^*, Z_i^{(1)}, Z^{(0)})$, by conditioning on the latent variables and parameters, the joint distribution for Y and D is degenerated, since we know the observed responses with certainty. Finally, the augmented posterior can be written as:

$$\begin{aligned} p(\Gamma, D^*, Z_i^{(1)}, Z^{(0)}|Y, D) = p(\Gamma)(& \prod_{i:D_i=1} \Pr(Y_i^{(1)} = j, D_i = 1|\Gamma) \times [I(D_i^* > 0) I(\tau_j^{(1)} < Z_i^{(1)} \leq \tau_{j+1}^{(1)})] \\ & + \prod_{i:D_i=0} \Pr(Y_i^{(0)} = j, D_i = 0|\Gamma) \times [I(D_i^* \leq 0) I(\tau_j^{(0)} < Z_i^{(0)} \leq \tau_{j+1}^{(0)})]). \end{aligned}$$

[12] gives the step-by-step algorithm on how to simulate the augmented posteriors from the above model settings.

4 Causal Effect for Ordinal Outcome

The causal inference using potential outcomes framework was proposed by [3]. This model is frequently referred to as Rubin Causal Model [1] . Rubin himself described the motivation and the history details of his model in [13]. The choice of estimand or the object of interest when the outcomes are continuous is the average treatment effect [3]. This quantity can be estimated in both randomized experiment and observational studies. For the valid causal inference from observed data, [14] reviewed various methods for causal inference from observational studies. These methods include propensity score, marginal structural models, and instrumental variables. The main idea for these methods is to control for the confounding effects.

However, in the ordinal outcomes, the average treatment effect is not well-defined due to the lack of scale or magnitude of the outcomes. The only information about the two consecutive outcomes is that the higher orders or categories are greater than the lower orders or categories. There are several researchers who have proposed the estimands regarding the ordinal outcomes. [15] proposed the difference in the distribution under the control and treatment. [16] proposed the conditional median under the monotonic treatment effects assumption as an estimand for causal inference of ordinal outcomes. The former estimand carries little information in some cases. The latter is more general and useful in practice.

To perform causal inference under potential outcomes framework, we assume the stable unit treatment value assumption [17]. Under this assumption, there will be only one version of the treatment and no interference among units. The pair $(Y_i^{(1)}, Y_i^{(0)})$ is defined as the potential outcomes of unit i under treatment and control, respectively. Let p_{kl} be the joint distribution of units whose potential outcome is k under treatment and l under control. Thus,

$$p_{kl} = \Pr(Y_i^{(1)} = k, Y_i^{(0)} = l), \quad k, l, = 1, 2, ..., J. \quad (4.1)$$

The $J \times J$ probability matrix $\mathbf{P} = [p_{kl}]$ characterize the joint distribution of the potential outcomes. The row and column sums of matrix $\mathbf{P}$ can be interpreted as elements in marginal distributions of potential outcomes under treatment and control, respectively. Let $p_{k+} = \sum_{l'=1}^{J} p_{kl'}$ and $p_{+l} = \sum_{k'=1}^{J} p_{k'l}$ be the column and row sums. Thus, the vectors $\mathbf{p}_1 = (p_{1+}, ..., p_{J+})$ and $\mathbf{p}_0 = (p_{+1}, ..., p_{+J})$ characterize marginal distribution of the potential outcomes under treatment and control, respectively.

From the marginal distributions defined above, we can define the distributional causal effects [18] for ordinal outcomes as:

$$\delta_j = \Pr(Y_i^{(1)} \geq j) - \Pr(Y_i^{(0)} \geq j) = \sum_{k \geq j} p_{k+} - \sum_{l \geq j} p_{+l}. \tag{4.2}$$

δ_j can be used to measure the difference between the marginal distributions of ordinal potential outcomes at each order j. [19] discussed the situation where δ_j have different signs for each order j. This situation led to the difficulty to decide whether the treatment or the control is preferable. [19] also proposed two causal effects that can be used to measure the beneficiary and strictly beneficiary of treatments for all units as follows:

$$\tau = \Pr(Y_i^{(1)} \geq Y_i^{(0)}) = \sum \sum_{k \geq j} p_{kl}, \tag{4.3}$$

and

$$\eta = \Pr(Y_i^{(1)} > Y_i^{(0)}) = \sum \sum_{k > j} p_{kl}. \tag{4.4}$$

The causal effect parameters proposed by [19] involve the association between ordinal potential outcomes of treatment and control. However, from the fundamental problem of causal effect estimation, we can never jointly observe both of them. Thus, the causal effects defined as τ and η are not identifiable. We propose the following procedure for causal inference:

1. Specify the parametric model discussed in Section 2 by selecting copula and marginal distribution functions. The model selection can be performed by using Bayesian model selection criteria such as BIC (Bayesian Information Criteria).
2. Estimate the parametric models specified in the first step by using Bayesian estimation discussed in Section 3 from observational data. Our model formulation can be used to correct the self-selection bias from the observational data. Thus, we can use the causal effect parameters discussed above under the assumption of the stable unit treatment value.
3. From the best model according to the BIC, use this model to impute the missing potential outcomes.
4. Estimate causal effect on the observed scale using the imputed data and observed data.

5 Case Study

In this section, we consider data from the Thai National Health Examination Survey, No.4 (NHES IV) data of 2009 as a case study [2]. Each of 19,948 individuals were classified into two groups based on their alcohol consumption behaviour. We considered these two groups: alcohol consumption and non-alcohol consumption as the treatment

and control groups, respectively. The ordinal outcomes of interest were blood pressure levels. The possible level of blood pressure were "normal level", "pre-hypertension level", and "hypertension level". The data are presented in Table 1.

The marginal distribution for each potential outcome model is specified as standard normal distribution. For the copula function, three models were estimated using the Independence copula, the Normal copula, and the Frank copula. We selected the best fitted model based on BIC, which is the Frank copula model. The covariates used in the model were individuals' characteristics such as gender, age, income, number of chronic diseases, indicators for occupations, and body-mass index.

For the choice of prior, we employed the independent prior for $p(\Gamma)$. In this paper, the model for marginal distributions is normal distribution. A natural prior $p(\beta)$ is independent normal distribution with means zero and large variances $(1000I_K)$, where K is the number of covariates and I_K is K-dimensional identity matrix. The more complicated prior choices are for the threshold α and copula dependence parameter θ. Notice that the threshold parameters are critical for model identifications, however, a default prior for the threshold parameter does not exist. Thus, we employed a "flat" prior by using a product of mean zero normal distribution with large variance.

We used the Bayesian approach that employ the data augmented as discussed in Section 3 to obtain the augmented posterior. We draw 10,000 posterior predictive samples and discard the first 2,000 draws as burn-in period. According to the lagged autocorrelation plots (results are not shown here) for all parameters, the results suggested that posterior distribution can be approximated reasonably by a moderate number of simulations.

From the best fitted model, we impute the missing potential outcomes and calculate the 3×3 joint probability matrix $\mathbf{P}$ of potential outcomes as discussed in Section 4 as

$$\boldsymbol{P} = \begin{bmatrix} 0.302 & 0.146 & 0.036 \\ 0.154 & 0.145 & 0.049 \\ 0.066 & 0.102 & 0.000 \end{bmatrix}$$

In our case study, we have $\tau = 0.769$ and $\eta = 0.322$. These can be interpreted as 76.9% of the population would induce higher or maintain their blood pressure levels, and 32.2% of the population would induce higher blood pressure levels if they consume alcohol.

Table 1: The distributions of individual blood pressure according to alcohol consumption status

Group	"Normal blood pressure"	"Prehypertension level"	"Having hypertension"
Control	5,151	4,236	3,342
Treatment	3,075	2,451	1,693

Acknowledgements : We acknowledge the partial support of the Center of Excellence in Econometrics, Faculty of Economics, Chiang Mai University, Thailand. This work was also supported in part by the Thailand Research Fund grant MRG-5980209.

References

[1] D.B. Rubin, Bayesian inference for causal effects: The role of randomization. The Annals of Statistics (1978) 34-58.

[2] K. Suknark, J. Sirisrisakulchai, S. Sriboonchitta, Reinvestigating the Effect of Alcohol Consumption on Hypertension Disease, Causal Inference in Econometrics (2016) 307-318.

[3] D.B. Rubin, Estimating causal effects of treatments in randomized and nonrandomized studies, Journal of educational Psychology 66 (5) (1974) 688-710.

[4] R.J. Little, D.B. Rubin, Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. Annu. Rev. Public Health 21 (2000) 121-145.

[5] P.R. Rosenbaum, D.B. Rubin, Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. Journal of the Royal Statistical Society, Series B (Methodological) (1983) 212-218.

[6] S.L. Morgan, W. Christopher, Counterfactuals and Causal Inference: Methods and Principles for Social Research, New York: Cambridge University Press, 2007.

[7] G.S. Maddala, Limited-Dependent and Qualitative Variables in Econometrics, Cambridge, Cambridge University Press, 1983.

[8] R.B. Nelsen, An Introduction to Copulas, second ed. Springer, New York, 2006.

[9] G. Casella, E.I. George, Explaining the Gibbs Sampler, The American Statistician 46 (1992) 167-174.

[10] S. Chib, E. Greenberg, Understanding the Metropolis-Hastings Algorithm, The American Statistician 49 (1995) 327-335.

[11] J. Albert, S. Chib, Bayesian analysis of binary and polychotomous response data, Journal of the American Statistical Association 88 (1993) 669-679.

[12] M. Li, J. Tobias, Bayesian Analysis of Treatment Effects in an Ordered Potential Outcomes Model, Advances in Econometrics, Modelling and Evaluating Treatment Effects in Econometrics (D.L. Millimet, J.A. Smith, E. Vytlacil, eds.) 21 (2008) 57-91.

[13] D.B. Rubin, Causal inference using potential outcomes, Journal of the American Statistical Association 100 (469) (2005) DOI 10.1198/016214504000001880.

[14] W. Martin, Making valid causal inferences from observational data, Preventive Veterinary Medicine 113 (2014) 281-297.

[15] S. Boes, Nonparametric analysis of treatment effects in ordered response models. Empirical Economics 44 (1) (2013) 81-109.

[16] A. Volfovsky, E.M. Airoldi, D.B. Rubin, Causal inference for ordinal outcomes, (2015) arXiv:1501.01234.

[17] D.B. Rubin, Randomization analysis of experimental data: The fisher randomization test comment, Journal of the American Statistical Association (1980) 591593.

[18] C. Ju, Z. Geng, Criteria for surrogate end points based on causal distributions, J.R. Stat. Soc. Ser. B Stat. Methodol. 72 (2010) 129-142.

[19] J. Lu, P. Ding, T. Dasgupta, Sharp Bounds of Causal Effects on Ordinal Outcomes, (2015) arXiv:1507.01542.

(Received 5 August 2016)
(Accepted 15 October 2016)