STATISTICS IN MEDICINE
Statist. Med. **17**, 957–970 (1998)

MODELLING ORDINAL RESPONSES FROM CO-TWIN CONTROL STUDIES

FRANK B. HU[1]*, JACK GOLDBERG[2], DONALD HEDEKER[3] AND WILLIAM G. HENDERSON[4]

[1] *Department of Nutrition, Harvard School of Public Health, 665 Huntington Ave., Boston MA 02115, U.S.A.*
[2] *Vietnam Era Twin Registry and Epidemiology and Biostatistics Program, School of Public Health, The University of Illinois at Chicago, U.S.A.*
[3] *Epidemiology and Biostatistics Program & Prevention Research Center, School of Public Health, The University of Illinois at Chicago, U.S.A.*
[4] *Vietnam Era Twin Registry and Cooperative Studies Program Coordinating Center, Hines, IL 60141, U.S.A.*

SUMMARY

The co-twin control design has been widely used in studying the effects of environmental factors on the development of diseases. For binary outcomes that arise from co-twin control studies, the conditional likelihood method is commonly used. This approach, however, does not readily extend to ordinal response data because the standard conditional likelihood does not exist for cumulative logit or proportional odds models. In this paper, we investigate the applicability of the random-effects and GEE approaches in analysing ordinal response data from co-twin control studies. Using both approaches, we re-analyse data from a co-twin control study of the impact of military services during the Vietnam era on post-traumatic stress disorders (PTSD). The ordinal models have considerably increased power in detecting the effects of exposure when compared to the analyses using a dichotomized response. We discuss the interpretation of the estimates from GEE and random-effects models in the context of the twin data. © 1998 John Wiley & Sons, Ltd.

1. INTRODUCTION

The co-twin control design is a powerful research methodology for studying the effects of environmental risk factors on the development of disease.[1] The research design may take either of two forms: cohort or case-control. In a cohort co-twin control study, exposure discordant twins are followed for the development of disease. In the case-control study, disease discordant pairs are

* Correspondence to: Frank B. Hu, Department of Nutrition, Harvard School of Public Health, 665 Huntington Avenue, Boston, MA 02115, U.S.A.

Contract grant sponsor: Department of Veterans Affairs Health Services Research and Development Services
Contract grant number: CSP#256

Contract grant sponsor: NIDA
Contract grant number: 1 RO1 DAO 4604-01

Contract grant sponsor: Great Lakes Veterans Affairs Health Services Research and Development Program
Contract grant number: LIP 41-065

CCC 0277–6715/98/090957–14$17.50

Received March 1996
Revised April 1997

examined for differences in antecedent exposures. While the co-twin control research design applies to either monozygotic twins (MZ) or dizygotic pairs (DZ), the design is most often used with MZ pairs, because MZ twins are genetically identical and discordance within pairs must be environmental in origin. The advantages of co-twin control design over other epidemiologic designs are twofold: (i) it provides an ideal control group, since MZ twins have the same genetic material, and they may also share a common rearing environment during their childhood and adolescent years, thus MZ pairs are perfectly matched on a multitude of known (genetics, age) and unknown potential confounding factors; and (ii) the twin design reduces the likelihood of selection bias if twins are ascertained from a twin registry.

For dichotomous dependent variables that arise from twin data, the conditional logistic regression is readily adapted to co-twin control studies.[2] Little work has been done on modelling ordinal responses for pair-matched data such as those that arise from co-twin control studies. The primary reason is that the standard conditional likelihood does not exist for the proportional odds model.[3–5] However, models have been developed recently for correlated ordinal data that arise from longitudinal and nested design studies. These models can be grouped into two classes:[6,7] one is termed the cluster-specific (CS) approach, such as the random-effects ordinal logistic model proposed by Hedeker and Gibbons;[8] the other is termed the population-averaged (PA) approach, such as, the generalized estimation equation (GEE) ordinal models developed by Lipsitz *et al.*[9] and Genge *et al.*[10]

In this paper, we investigate the applicability of the random-effects and GEE approaches in analysing ordinal response data from co-twin control studies. Using both approaches, we re-analyse data from a co-twin control study of the impact of military service during the Vietnam era on post-traumatic stress disorder (PTSD).[11] In particular, we compare the random-effects approach to the GEE approach for ordinal response data. We also compare analyses that respect the ordinal nature of the response to a dichotomized analysis.

The remainder of this paper is organized as follows. Section 2 describes the data set from the Vietnam Era Twin Registry. Section 3.1 briefly describes the conditional likelihood approach for binary outcomes that arise from co-twin control studies. Section 3.2 describes the GEE approach for ordinal outcomes. Section 3.3 discusses the random-effects approach for ordinal responses. Section 4 presents results from the random-effects and GEE models by analysing the dichotomized and the original ordinal response data. Note that we do not present results from the conditional logistic models applied to the dichotomized responses because they have been presented elsewhere.[11] Section 5 discusses the results.

2. THE DATA SET: SERVICE IN SOUTHEAST ASIAN (SEA) AND PTSD IN VIETNAM VETERAN TWIN PAIRS

The data of interest were collected in a co-twin control study of the effects of the Vietnam war on post-traumatic stress disorder (PTSD). The sources of data, classification of zygosity, and measurements of PTSD were described in detail by Goldberg *et al.*[11] In summary, the data used in the study were derived from the Vietnam Era (1965 to 1975) Twin Registry. Information regarding Registry members was obtained in 1987 with completion of a survey questionnaire by mail or telephone or via in-person interviews (74·4 per cent response rate). Zygosity was determined using a previously validated questionnaire technique supplemented with blood group-typing data abstracted from military records.

Information regarding the demographic and military service experience of veterans was obtained from the military service record and the health survey. The exposure variable, service in southeast Asia (SEA), was derived from the health survey based on the response to the following question: 'When you were in the military, were you stationed in Vietnam, Laos or Cambodia; in the waters in or around these countries; or fly in missions over these areas?'. The identification of PTSD symptoms was based on 12 items (many specific to military services experience) in the health survey that closely resembled the list of the symptoms included in DSM-IIIR manual.[12] For each item, every respondent was asked to indicate the frequency of the symptom (very often, often, sometimes, almost never, or never) during the preceding 6 months. In addition, information on important confounding variables was also collected, such as age at the time of enlistment, education level, and branch of service (army, marine, air force and navy).

In previous analyses,[11] the five-level ordinal scale was dichotomized by considering any individual who indicated a frequency of sometimes, often, or very often as a positive response for that symptom. Strong associations between SEA services and the dichotomous measures of PTSD symptoms were obtained using the conditional logistic regression model for matched pairs. In this paper, we re-analyse only one of the PTSD symptoms: nightmares about military experiences.

After eliminating missing data on the response variable, a total of 2497 monozygotic pairs were eligible for the present analyses. Of these pairs, 500 both served in SEA, 854 were discordant for SEA service (one in SEA and the other not in SEA), and for 1143 pairs neither served in SEA. In the group of twins where neither member of the pair served in SEA, the prevalence of nightmares was 6·1 per cent. By contrast, among the twins concordant for service in SEA, the prevalence of nightmares was much greater (23·7 per cent). Among the 854 SEA service discordant pairs (1708 individuals), the prevalence of having nightmares in siblings who did not serve in SEA was similar to that of concordant non-SEA pairs, 8·6 per cent, while the prevalence among siblings who served in SEA was similar to that of pairs concordant for SEA service, 26·5 per cent.

3. METHODS

3.1. Conditional likelihood approach for binary response data from twin studies

Suppose that we wish to study the effects of exposure on a dichotomous outcome. A logistic model can be written as follows:

$$\Pr(Y_{ij} = 1 | X_{ij}, i) = \frac{\exp(\tau_i + X'_{ij}\beta)}{1 + \exp(\tau_i + X'_{ij}\beta)} \tag{1}$$

where Y_{ij} represents the binary outcome variable for twin j ($j = 1, 2$) in pair i ($i = 1, \ldots, n$). X_{ij} is a $p \times 1$ vector of subject-level covariates for twin j in pair i. τ_i represents effects for the variation among twins.

To obtain a consistent estimate of β, we can apply conditional likelihood methods.[2] The conditional likelihood is

$$\prod_i (1 + \exp(-D'_i\beta))^{-1} \tag{2}$$

where $D_i = X_{i2} - X_{i1}$ is the difference in the two covariate vectors for the twins in pair I. Note that the product is over the n pairs of twins.

The odds ratio of event comparing two twins with and without exposure in pair I:

$$\psi_i = \frac{\Pr(Y_{2i}=1|X_{2i})/\Pr(Y_{2i}=0|X_{2i})}{\Pr(Y_{1i}=1|X_{1i})/\Pr(Y_{1i}=0|X_{1i})} = \exp(D_i'\beta). \tag{3}$$

Through the conditional likelihood approach, we remove the α_i parameters in equation (1) by conditioning on their sufficient statistics in the likelihood. In addition, this method only uses data from pairs that are discordant on both response and the covariates. Thus, the concordant pairs contribute no information to the likelihood. For a single binary exposure variable, the estimated odds ratio reduces to the ratio of discordant pairs.[13] Since conditional likelihood is unaffected by the sampling scheme (that is, retrospective versus prospective sampling),[14] the method can be used in twin studies of either case-control or cohort forms.

Unfortunately, the conditional likelihood approach does not extend readily to ordinal response data since reduced sufficient statistics do not exist for the cumulative logistic model (proportional odds model).[3–5] Conaway[15] has applied the conditional likelihood approach using the Rasch model to analyse repeated categorical measurements. However, Conaway's model is not a proportional odds model. Agresti and Lang[4] have proposed a method for approximating the conditional likelihood in a cumulative logit model for repeated ordinal data by fitting conditional Rasch models simultaneously for all binary recodings of the ordinal response. However, the model is a log-linear Rasch model and only allows for limited stratification.

Since we cannot derive the standard conditional likelihood for the proportional odds model, we investigate the possibility of employing other methods to fit proportional odds models for ordinal response data that arise from co-twin control studies. As discussed above, our primary interest is in the GEE and random-effects ordinal models.

3.2. The GEE approach for ordinal response data

Several authors have extended the original GEE models proposed by Liang and Zeger[16] to correlated ordinal outcomes.[9,10] For the twin data, a GEE ordinal model can be written as

$$\log\left[\frac{P(Y_{ij}\leqslant k)}{1-P(Y_{ij}\leqslant k)}\right] = \alpha_k - X_{ij}'\beta \qquad k=1,\ldots,K-1 \tag{4}$$

where $k=1,\ldots,K-1$ for the K categories of the ordinal outcome variable Y, and α_k is a vector of $K-1$ intercepts.

This model is an extension of the proportional odds model for independent ordinal data described in McCullagh,[17] Agresti[18] and elsewhere. As with the proportional odds models for independent data, the left-hand side of the equation (1) specifies $K-1$ cumulative logits, each contrasting the combined first k categories to the remaining combined $(K-k)$ categories. For example, with four possible response categories (coded as 1, 2, 3, 4), the following three cumulative logistics are indicated by the model:

$$\log\left[\frac{P(Y_{ij}\leqslant 1)}{1-P(Y_{ij}\leqslant 1)}\right] = \log\left[\frac{P(Y_{ij}=1)}{P(Y_{ij}=2,3,4)}\right] \tag{5}$$

$$\log\left[\frac{P(Y_{ij}\leqslant 2)}{1-P(Y_{ij}\leqslant 2)}\right] = \log\left[\frac{P(Y_{ij}=1,2)}{P(Y_{ij}=3,4)}\right] \tag{6}$$

$$\log\left[\frac{P(Y_{ij}\leqslant 3)}{1-P(Y_{ij}\leqslant 3)}\right] = \log\left[\frac{P(Y_{ij}=1,2,3)}{P(Y_{ij}=4)}\right]. \tag{7}$$

The proportional odds model assumes that the effect of the exposure variable is the same across these $K-1$ cumulative logits, or proportional across the cumulative odds. Thus, a single effect is estimated for exposure, which is the simultaneous effect of exposure on the $K-1$ cumulative logits. The odds of a response in a category greater than k (for any fixed k) is multiplied by $\exp(\beta)$ for every unit change in the exposure variable.

Lipsitz *et al.*[9] and Barnwell *et al.*[19] have described the estimation procedure in detail. Let $y_{ijk}=1$ if the jth twin for the ith pair is in response category k, and zero otherwise; y_{ij} is a vector of the binary response indicators for twin j $[y_{ij1}, \ldots, y_{ij(K-1)}]$. Combining the response vectors yields $y_i=[y'_{i1}, y'_{i2}]'$. The GEE estimate of β is the solution to the following estimating equation:

$$U(\beta)=\sum_{i=1}^{n}\frac{\partial\pi'_i}{\partial\beta}\,V_i(\rho)^{-1}(y_i-\pi_i)=0 \tag{8}$$

where $\pi_i=E(y_i)$ is the mean response vector for pair i, $(\partial\pi'_i/\partial\beta)$ is the vector of first partial derivatives of the mean response π_i with respect to the regression coefficients β, and $V_i(\rho)=\mathrm{cov}(y_i;\pi_i,\rho)$ is the working covariance matrix.

Under this approach, we use the 'working' correlation matrix for the vector of repeated observations to account for the within-pair dependency. For a binary outcome, we can use a simple correlation matrix for the correlation between twins:

$$R(\rho)=\begin{bmatrix}1 & \rho_{21}\\ \rho_{12} & 1\end{bmatrix} \tag{9}$$

where ρ_{12} is the correlation of the response between the twins, and $\rho_{12}=\rho_{21}$. In the ordinal case, since the marginal distribution of the response is multinomial, the correlation matrix takes a multivariate form:[10]

$$R(\rho)=\begin{bmatrix}I_{K-1} & R^*_\rho\\ R^{*\prime}_\rho & I_{K-1}\end{bmatrix} \tag{10}$$

where I_{k-1} is a $(K-1)\times(K-1)$ identity matrix and R^*_ρ is a $(K-1)\times(K-1)$ correlation matrix of the $K-1$ binary response indicators between the twins. For example, when $K=4$, R^*_ρ is a 3×3 symmetric matrix:

$$R^*_\rho=\begin{bmatrix}\rho_{11} & & \\ \rho_{21} & \rho_{22} & \\ \rho_{31} & \rho_{32} & \rho_{33}\end{bmatrix}. \tag{11}$$

The GEE solution involves an iterative procedure that alternates between quasi-likelihood methods for estimating β and a robust method for estimating ρ as a function of the parameter estimates until convergence occurs. An important feature of the GEE is that it gives consistent estimators of the regression coefficients and of their variance in large samples even if the assumed correlation matrix is misspecified. Very recently, Barnwell *et al.*[19] have implemented the Lipsitz *et al.* procedure in the software SUDAAN.

Although the GEE approach can account for the correlation between the twins through robust variance estimation, it models the marginal distributions of the response and treats the correlated data as though it were unpaired. So the analysis is analogous to an unmatched analysis.

3.3. A random-effects logistic model for ordinal response data

An alternative approach for correlated ordinal response data is the random-effects or mixed-effects approach. For the twin data, we can write an ordinal logistic model as follows:

$$\log\left[\frac{P(Y_{ij} \leqslant k)}{1 - P(Y_{ij} \leqslant k)}\right] = \alpha_k - (X'_{ij}\beta + \upsilon_i) \tag{12}$$

where υ_i is the random effect for twin pair i and assumed to be distributed as $\mathrm{N}(0, \sigma_\upsilon^2)$.

With the proportional odds assumption, the conditional probability of response occurring in category k for twin j within a given pair i is

$$P(Y_{ij} = k \,|\, \upsilon_i, \beta) = \Psi(z_{ijk}) - \Psi(z_{ij(k-1)}) \tag{13}$$

where

$$z_{ijk} = \alpha_k - (X'_{ij}\beta + \upsilon_i)' \tag{14}$$

$$\Psi(z_{ijk}) = 1/[1 + \exp(-z_{ijk})] \tag{15}$$

and $\Psi(0) = 0$ and $\Psi(K) = 1$.

Let Y_i denote the two responses from pair i. The probability of any pattern Y_i given β and υ_i is equal to the product of the probabilities of the two responses:

$$\ell(Y_i \,|\, \beta, \upsilon_i) = \prod_{j=1}^{2} \prod_{k=1}^{K} [\Psi(z_{ijk}) - \Psi(z_{ij(k-1)}]^{d_{ijk}} \tag{16}$$

where $d_{ijk} = 1$ if $Y_{ij} = k$, and $d_{ijk} = 0$ if $Y_{ij} \neq k$.

The marginal probability for the response Y_i in the population is

$$h(Y_i) = \int_\upsilon \ell(Y_i \,|\, \beta, \upsilon_i) g(\upsilon)\, \mathrm{d}(\upsilon) \tag{17}$$

where $g(\upsilon)$ represents the population distribution of the random effect, υ. For integration over υ, one can use Gauss–Hermite quadrature method to perform the integration numerically by summing over a specified number of quadrature points. One can use an iterative Fisher scoring solution to obtain model parameters and standard errors.[8]

The random effect υ_i reflects the dependency between two twins within pair i, over and above the influence of the other model terms. This dependency may be due to unmeasured shared genetic and environmental factors. As the correlation of responses between MZ twins is high, values of υ_i will deviate from zero, resulting in a population variance σ_υ^2 that is greater than zero. Since the standard logistic distribution has a variance equal to $\pi^2/3$, we can express the intracluster correlation (ICC) for MZ twins as

$$\mathrm{ICC} = \frac{\sigma_\upsilon^2}{(\sigma_\upsilon^2 + \pi^2/3)}. \tag{18}$$

Notice that we can consider the random-effects logistic model for a dichotomous response as a special case of the ordinal model described in equation (12). With a dichotomous response variable, the random-effects model reduces to

$$\log\left[\frac{P(Y_{ij}=1)}{1-P(Y_{ij}=1)}\right]=\beta_0+X'_{ij}\beta+\upsilon_i. \tag{19}$$

The coefficient β measures the change of the logit of the probability of response between one twin without the exposure and the other with the exposure for a given pair (that is, controlling for υ_i). Hence, the odds ratio e^{β} describes the within-pair change in the risk of developing the outcome. Clearly, this interpretation is the same as the conditional likelihood model. Additionally, Neuhaus *et al.*[20] have shown that for subject-level covariates that have no between-cluster component, random-effects logistic models and conditional likelihood logistic models yield nearly identical estimates. Therefore, we expect that the two approaches for modelling dichotomized binary responses provide the same results when only exposure discordant pairs are used in the analyses.

Use of the random-effects approach in modelling co-twin control data is intuitively appealing. We can think of the responses from each member of an MZ twin pair as repeated measures for one pair. The random-effects model assumes that the logit varies from one pair to the next by υ_i. This assumption is reasonable because each pair of twins has its own unique shared genetic and environmental background. Thus, this variability reflects natural heterogeneity due to unmeasured genetic and environmental factors among all the twin pairs. This heterogeneity is represented by a Gausssian probability distribution, that is, the random pair effect, υ_i, which is assumed normally distributed in the population.

We fit the random-effects logistic model in this paper using a FORTRAN program called MIXOR written by Hedeker and Gibbons.[21] MIXOR can fit both dichotomous and ordinal responses and comprises both logistic and probit response functions. Since there is no closed-form solution to the likelihood equation in the random-effect model, we used Gauss–Hermite quadrature to perform the numerical integration. Usually, the more points used, the more accurate the approximation, however the more time it takes. In the present paper, we used 20 quadrature points for model estimation.

4. RESULTS

Table I shows the distribution of 854 exposure discordant pairs by the frequency of reported nightmares about military service in the past six months. Clearly, there were more data in the upper off-diagonal than in the lower off-diagonal, suggesting that in general, the twins who served in SEA were more likely to report nightmares than those who did not serve in SEA.

Table II shows parameter estimates, standard errors, and Wald χ^2 obtained from random-effects and GEE logistic models for exposure discordant pairs ($n = 854$). We fit the models for both a dichotomized response (grouping sometimes, often, and very often as a positive response) and the five-level ordinal response. For both dichotomized and ordinal responses, we estimated two models, one with only the exposure variable, service in SEA (unadjusted), and the other including potential person-varying confounding variables, such as education, branch of service, and the pair-varying variable, age (adjusted). Consider first the results from the random-effects models. For the dichotomized response, the estimate and standard error of the exposure (service

Table I. The distribution of exposure discordant pairs by the reported frequency of nightmares

Not in SEA	Service in SEA					
	Never	Almost never	Sometimes	Often	Very Often	Total
Never	391	140	113	20	8	672
Almost never	40	24	32	8	5	109
Sometimes	18	11	26	3	4	62
Often	2	1	2	1	1	7
Very often	0	1	3	0	0	4
Total	451	177	176	32	18	854

in SEA) were nearly identical to those obtained from conditional logistic models by Goldberg *et al.*[11] Adjusting for education level, branch of service, and age at the time of enlistment had no impact on the effect of the exposure variable. The adjusted odds ratio for SEA service and nightmares was 5·70 (95 per cent confidence interval: 3·92–8·27). This odds ratio can be interpreted as a twin exposed to SEA was 5·70 times more likely to have nightmares than his co-twin who was not exposed. The estimated intracluster correlation (ICC) was around 0·40 (not shown in the table), suggesting moderately high correlation in responses between the MZ twins. The confounding variables had little effect on the estimates of the random-effect variance term, σ_υ and the ICC.

Both the estimates and standard errors from the GEE models were considerably smaller than those from the random-effects models. The crude odds ratio of nightmare for the exposure was 3·86 (95 per cent CI: 2·99–4·98). In contrast to that from the random-effects approach, this odds ratio can be interpreted as that subjects exposed to SEA were 3·86 times more likely to have nightmares than those not exposed. Adjustment for the confounding variables had little impact on the association.

In the random-effects models with ordinal responses, the parameter estimates for SEA were slightly smaller than those obtained from the corresponding models with dichotomized responses. However, the estimated standard errors were considerably smaller than those from the models with dichotomized responses. Thus, the test statistics (Wald χ^2) for the exposure variable obtained from the ordinal models were considerably larger than those from the logistic models with dichotomized responses, indicating increased power for the ordinal model. The estimated σ_υ in the ordinal models were slightly reduced when compared to corresponding models with dichotomized responses. Again, adjustment for confounding variables had little impact on the effects of the exposure, and the estimates of σ_υ.

Compared to the GEE models with dichotomized responses, the estimates and standard errors of the exposure from the ordinal GEE model were also reduced, but the values of Wald χ^2 increased. As with the dichotomous case, the GEE estimates from the ordinal models were smaller than those from the corresponding random-effects models.

In both GEE and random-effects models, it is possible to include not only exposure discordant pairs ($n = 854$ pairs), but also exposure concordant pairs ($n = 1643$ pairs) in the analyses. Table III lists the parameter estimates and standard errors from random-effects and GEE logistic models for all pairs. Again, the models were estimated for both dichotomized and the original ordinal responses. In all models, the estimates of the exposure variable were noticeably greater

Table II. The association of SEA service and nightmares: parameter estimates, standard errors and Wald χ^2 from random-effects logistic models for exposure discordant pairs ($n = 854$)

Parameters	Dichotomized response				Ordinal response			
	Random-effects		GEE		Random-effects		GEE	
	Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted
α_1	3·14	− 2·50	2·37	− 2·06	1·73	− 3·09	1·32	− 2·41
	(0·24)	(1·37)	(0·12)	(1·20)	(0·12)	(1·08)	(0·08)	(0·94)
	(175·56)	(3·31)	(390·06)	(2·95)	(193·2)	(8·24)	(250·27)	(6·60)
α_2					2·97	− 1·84	2·28	− 1·44
					(0·16)	(1·07)	(0·10)	(0·94)
					(339·66)	(2·96)	(559·80)	(2·34)
α_3					5·11	0·32	4·07	0·39
					(0·24)	(1·05)	(0·15)	(0·95)
					(441·00)	(0·009)	(741·47)	(0·17)
α_4					6·25	1·47	5·12	1·45
					(0·32)	(1·04)	(0·23)	(0·99)
					(382·59)	(2·02)	(516·65)	(2·16)
Service in SEA	1·73	1·74	1·35	1·38	1·60	1·60	1·23	1·25
	(0·19)	(0·19)	(0·13)	(0·13)	(0·13)	(0·13)	(0·09)	(0·09)
	(83·72)	(80·82)	(105·24)	(104·50)	(149·08)	(147·87)	(177·96)	(173·98)
Education		− 0·33		− 0·25		− 0·17		− 0·10
		(0·15)		(0·12)		(0·12)		(0·10)
		(4·80)		(4·07)		(2·05)		(0·94)
Branch of service (Army as reference):								
Marine		0·59		0·40		0·48		0·38
		(0·31)		(0·23)		(0·24)		(0·20)
		(3·65)		(3·02)		(3·88)		(3·65)
Air force		− 0·69		− 0·57		− 0·59		− 0·47
		(0·27)		(0·23)		(0·20)		(0·16)
		(6·40)		(6·14)		(8·70)		(8·24)
Navy		− 0·54		− 0·45		− 0·52		− 0·42
		(0·25)		(0·19)		(0·18)		(0·15)
		(4·67)		(5·78)		(7·90)		(7·93)
Age		− 0·12		− 0·09		− 0·11		− 0·09
		(0·03)		(0·03)		(0·03)		(0·02)
		(11·49)		(9·00)		(15·68)		(12·19)
σ_v	1·48	1·40			1·34	1·30		
	(0·21)	(0·22)			(0·14)	(0·14)		
	(48·72)	(42·38)			(96·63)	(87·67)		

than those obtained from corresponding models with exposure discordant pairs, while the standard errors were noticeably smaller. Again, the models with ordinal response variable were more efficient than those with the dichotomized response variable. With all pairs included in the analyses, the standard errors for all the confounding variables were reduced, and the test statistics increased. The estimates of σ_v also increased. Adjustment for confounding variables slightly increased the effects of the exposure. Also, GEE estimates were in general smaller than those from the random-effects models.

Table III. The association of SEA service and nightmares: parameter estimates, standard errors and Wald χ^2 from random-effects logistic models for all pairs ($n = 2497$)

Parameters	Dichotomized response				Ordinal response			
	Random-effects		GEE		Random-effects		GEE	
	Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted
α_1	3·68	− 1·97	2·61	− 1·52	1·97	− 2·56	1·44	− 1·83
	(0·16)	(0·86)	(0·08)	(0·65)	(0·08)	(0·62)	(0·05)	(0·47)
	(502·66)	(5·20)	(1179·4)	(5·47)	(635·04)	(17·14)	(876·16)	(15·05)
α_2					3·44	− 1·18	2·47	− 0·78
					(0·10)	(0·61)	(0·06)	(0·47)
					(1022·7)	(3·69)	(1694·7)	(2·76)
α_3					5·39	0·87	4·14	0·91
					(0·15)	(0·61)	(0·10)	(0·48)
					(1267·4)	(2·02)	(1714·0)	(3·53)
α_4					6·65	2·13	5·29	2·06
					(0·21)	(0·62)	(0·15)	(0·50)
					(1041·4)	(11·97)	(1246·1)	(16·97)
Service in SEA	2·03	2·19	1·50	1·66	1·75	1·88	1·30	1·42
	(0·13)	(0·14)	(0·09)	(0·10)	(0·09)	(0·09)	(0·06)	(0·07)
	(228·92)	(237·16)	(264·68)	(278·34)	(371·72)	(392·00)	(402·80)	(427·66)
Education		− 0·25		− 0·19		− 0·09		− 0·04
		(0·10)		(0·08)		(0·07)		(0·06)
		(6·66)		(5·85)		(1·48)		(0·44)
Branch of service (Army as reference):								
Marine		0·46		0·32		0·50		0·40
		(0·19)		(0·16)		(0·15)		(0·12)
		(5·62)		(4·00)		(10·34)		(10·18)
Air force		− 0·79		− 0·63		− 0·47		− 0·36
		(0·19)		(0·15)		(0·13)		(0·10)
		(17·31)		(17·97)		(13·62)		(12·96)
Navy		− 0·59		− 0·48		− 0·48		− 0·38
		(0·16)		(0·12)		(0·11)		(0·08)
		(14·59)		(16·80)		(18·66)		(19·45)
Age		− 0·12		− 0·09		− 0·11		− 0·08
		(0·02)		(0·02)		(0·02)		(0·01)
		(30·25)		(29·48)		(45·97)		(42·25)
σ_v	1·72	1·62			1·48	1·42		
	(0·14)	(0·13)			(0·08)	(0·08)		
	(160·02)	(146·41)			(332·33)	(306·60)		

5. DISCUSSION

The substantive conclusion drawn from this study is the same as in the original paper.[11] That is, service in Southeast Vietnam increased the risk of having the PTSD symptoms (nightmares about military experiences in this study). However, in this study, we observed considerably increased power in detecting exposure effects when compared to the analyses using a dichotomized response variable. For unmatched data, simulations have demonstrated that the asymptotic

relative efficiency of simple logistic regressions applied to dichotomized responses ranges from 30 per cent to 80 per cent when compared to ordinal regressions (proportional odds models).[22]

The GEE and random-effects models present two distinct methods for analysing correlated categorical outcomes. The GEE approach models the marginal distribution of repeated observations for the twins and is analogous to an unmatched analysis, whereas the random-effects approach models the joint distribution of responses for the twins and the parameter estimates have a 'pair-specific' interpretation. In the dichotomous case, the odds ratio from a GEE model is the odds of the event among subjects with the exposure divided by the odds among subjects without the exposure. Since this odds ratio is the ratio of sub-population risk, regardless of the pairing, they are referred as population-averaged effects. On the other hand, the odds ratio from a random-effects model describes the change in the risk of disease or outcome for a given pair whose exposure status changes within the pairs (that is, exposure discordant twin pairs). Hence, the interpretation of the random-effects odds ratio is the same as the conditional likelihood approach when only exposure discordant pairs are analyzed. From this perspective, the random-effects approach is more appropriate than the GEE approach in analysing co-twin control data. Nevertheless, there are some connections between the two approaches for both binary and ordinal models.[23] Specifically, the estimates from the two approaches approximately satisfy the following equation:

$$\beta_{\mathrm{GEE}} \simeq \frac{\beta_{\mathrm{RRM}}}{\sqrt{(1 + 0{\cdot}346\sigma_v^2)}}. \tag{20}$$

One important feature of the random-effects approach is that it can readily accommodate missing data in the response variables. As Laird[24] points out, random-effects models using maximum likelihood estimation provide valid inferences in the presence of ignorable non-response. By ignorable non-response, it is meant that the probability of non-response in one twin is dependent on the observed response of another twin within a pair and/or observed covariates.

The GEE approach discussed in this paper, however, requires stronger assumptioms regarding missing data, that is, missing completely at random (MCAR). In both random-effects and GEE approaches, missing data mechanisms may need to be specified in the model in the presence of non-ignorable missingness.[25] The conditional logistic model cannot accommodate any missing data. In this study, we did not have a chance to investigate the impact of missing data since only 2 per cent of the responses were missing.

In the random-effects model, when only discordant pairs are used in the analyses, the estimated coefficient for the exposure variable involves only within-pair comparisons. Since the twins in each pair are identical in genetic material and may share many unknown environmental factors, these types of comparisons are immune from many confounding effects that may occur in the usual epidemiologic studies. However, exclusive use of discordant pairs means that a larger proportion of the data are 'non-informative' and not included in estimation or testing. As discussed by Hrubec and Robinette,[1] when only discordant pairs are used, the sample may become so small that it may not have enough power to detect a difference in response. In the present study, it means that 66 per cent of the data provided no information on the effect of exposure. This study shows that the inclusion of both exposure discordant and concordant pairs in the analysis can substantially increase statistical power. However, the interpretation of the estimates is somewhat altered. In this situation, the estimated coefficient for the exposure variable involves both a within-pair comparison based on differences between twins for the same pair, and a between-pair comparison of average levels. Obviously, a between-pair comparison is not

immune from the confounding effects that may occur in typical epidemiologic studies. To address this problem, Ten Have *et al.*[26] suggested that in a random-effects model, this type of association could be decomposed into a within-cluster component and a between-cluster component. For example, to include all pairs in the analyses, we can rewrite equation (12) as

$$\log\left[\frac{P(Y_{ij} \leqslant k)}{1 - P(Y_{ij} \leqslant k)}\right] = \alpha_k - (\beta_B \bar{X}_i' + \beta_W (X_{ij} - \bar{X}_i)' + \upsilon_i) \tag{21}$$

where β_B represents the effect based on a between-pair comparison of the average level of exposure across all pairs, while β_W represents the effect based on a within-pair comparison only. The interpretation of β_B and β_W is much clearer than that of the overall effect. We reran the random-effects ordinal model including all pairs using this decomposition of the exposure variable (service in SEA) and obtained β_W which was slightly reduced compared to the overall estimate (1·71 versus 1·75). Note that this decomposed within-pair estimate (1·71) is larger than that obtained considering only exposure discordant pairs (1·60 from Table II) due to the larger estimate of σ_υ in the decomposed analysis.

In random-effects models, the effects of the individual-level confounding variables also involve both within-pair and between-pair comparisons. Thus, they may differ from the effects estimated through the conditional likelihood approach. In addition, we can estimate the effects of pair-level variables (for example, age) with use of the random-effects approach, but not with use of the conditional likelihood approach, though we can estimate the interactions between pair-level variables and the exposure using both approaches.

In this paper, we used the random-effects and GEE approaches to analyse ordinal response data from a co-twin control study. For both approaches, the ordinal models have considerably increased power in detecting the effects of exposure when compared to the analyses using a dichotomized response. The interpretation of the effects of the exposure, however, is not the same between the two approaches.

ACKNOWLEDGEMENTS

We are grateful to the referee for many helpful suggestions. This study was supported by the Department of Veterans Affairs Health Services Research and Development Services (CSP #256). Partial support was provided by the NIDA (Bethesda, MD) grant 1 RO1 DAO 4604-01 and the Great Lakes Veterans Affairs Health Services Research and Development Program, Ann Arbor, MI, LIP 41-065.

The authors acknowledge the work of the following people: (i) Midwest Center for Health Services and Policy Research: Vietnam Era Twin Registry, Director, W.G. Henderson, Ph.D., Epidemiologist, J. Goldberg, Ph.D.; Registry Programmer, K. Bukoeski; Co-ordinator, M. E. Vitek; (ii) VET Registry Advisory Committee: A. G . Bearn, M. D. (past), G. Chase, Sc. D. (past), T. Colton, Sc. D., W. E. Nance, M. D., Ph.D., R. S. Paffenbarger, Jr., M. D., Dr. P. H., M. M. Weissman, Ph.D., and R. R. Willams, M. D. (iii) DVA Chief Research & Development Officer: John R. Feussner, M. D.; DVA health Services Research & Development Filed and Core Support Programs: Acting Director, S. Meehan, M.B.A., Ph.D.; Assistant Director, C. Welch, III, Ph.D., Program Assistant, J. J. Gough.

The following organizations provided invaluable support in the conduct of this study: Department of Defense; National Personal Records Center, National Archives and Records Administration;

the Internal Revenue Service; National Opinion Research Center; National Research Council, National Academy of Sciences; the Institute for Survey Research, Temple University.

The authors also appreciate the assistance provided by the CSP # 256 co-chairmen, Drs. S. Eisen and W. True. More importantly, the authors gratefully acknowledge the continued co-operation and participation of the members of the Vietnam Era Twin Registry. Without their contribution this research would not have been possible.

REFERENCES

1. Hrubec, Z. and Robinette, C. D. 'The study of human twins in medical research', *New England Journal of Medicine*, **310**, 435–441 (1994).
2. Breslow, N. E. and Day, N. E. *Statistical Methods in Cancer Research. Volume* 1 – *The Analysis of Case-control Studies*, International Agency for Research on Cancer, IARC Scientific Publications No. 32. IARC, Lyon 1980.
3. McCullagh, P. 'A logistic model for paired comparisons with ordered categorical data', *Biometrika*, **64**, 449–453 (1977).
4. Agresti, A. and Lang, J. B. 'A proportional odds model with subject-specific effects for repeated ordered categorical responses', *Biometrika*, **80**, 527–534 (1993).
5. Ten Have, T. R., Landis, J. R. and Weaver, S. L. 'Association models for periodontal disease progression: A comparison of methods for clustered binary data', *Statistics in Medicine*, **14**, 423–429 (1995).
6. Neuhaus, J. M., Kalbleisch, J. D. and Hauck, W. W. 'A comparison of cluster-specific and population-averaged approaches for analysing correlated binary data', *International Statistical Review*, **59**, 25–36 (1991).
7. Diggle, P., Liang, K-Y. and Zeger, S. L. *Analysis of Longitudinal Data*, Clarendon Press, Oxford, 1994.
8. Hedeker, D. and Gibbons, R. D. 'A random-effects ordinal regression model for multilevel analysis', *Biometrics*, **50**, 933–944 (1994).
9. Lipsitz, S. R., Kim, K. and Zhao, L. 'Analysis of repeated categorical data using generalized estimating equations', *Statistics in Medicine*, **13**, 1149–1163 (1994).
10. Genge, S. J., Linton, K. L. P., Scott, D. L. and Klein, R. 'A comparison of methods for correlated ordinal measures with ophthalmic applications', *Statistics in Medicine*, **14**, 1961–1974 (1995).
11. Goldberg, J., True, W. R., Eisen, S. A. and Henderson, W. G. 'A twin study of the effects of the Vietnam war on posttraumatic stress disorder', *Journal of the American Medical Association*, **263**, 1227–1232 (1990).
12. *Diagnostic and Statistical Manual of Mental Disorders*, *Third Edition*, *Revised*, American Psychiatric Association, Washington DC, 1987.
13. McNemar, Q. 'Note on the sampling error of the difference between correlated proportions or percentages', *Psychometrika*, **12**, 153–157 (1947).
14. Neuhaus, J. M. and Jewell, N. P. 'The effect of retrospective sampling on binary regression models for clustered data', *Biometrics*, **46**, 977–990 (1990).
15. Conaway, M. R. 'Analysis of repeated categorical measurements with conditional likelihood methods, *Journal of the American Statistical Association*, **84**, 53–62 (1989).
16. Liang, K-Y. and Zeger, S. L. 'Longitudinal data analysis using generalized linear models', *Biometrika*, **73**, 13–22 (1986).
17. McCullagh, P. 'Regression models for ordinal data (with discussion)', *Journal of the Royal Statistical Society*, *Series B*, **42**, 109–142 (1980).
18. Agresti, A. 'Tutorial on modeling ordered categorical response data', *Psychological Bulletin*, **105**, 209–301 (1989).
19. Barnwell, B. G., Bieler, G. S. and Shah, B. V. 'SUDDAN Technical Report: The MULTILOG Procedure', Release 6.6. Research Triangle Institute, Research Triangle Park, NC, 1996.
20. Neuhaus, J. M., Kalbfleisch, J. D. and Hauck, W. W. 'Conditions for consistent estimation in mixed models for binary matched-pairs data', *Canadian Journal of Statistics*, **22**, 139–148 (1994).
21. Hedeker, D. and Gibbons, R. D. 'MIXOR: A computer program for mixed-effects ordinal probit and logistic regression analysis', *Computer Methods and Programs in Biomedicine*, **49**, 157–176 (1996).

22. Armstrong, B. G. and Sloan, M. 'Ordinal regression models for epidemiologic data', *American Journal of Epidemiology*, **129**, 191–204 (1989).
23. Ten Have, T. R., Landis, J. R. and Hartzel, J. 'Population-averaged and cluster-specific models for clustered ordinal response data', *Statistics in Medicine*. In press.
24. Laird, N. M. 'Missing data in longitudinal studies', *Statistics in Medicine*, **7**, 305–315 (1988).
25. Little, R. J. A. 'Modeling the drop-out mechanism in repeated-measures studies', *Journal of American Statistical Association*, **90**, 1112–1121 (1995).
26. Ten Have, T. R., Landis, J. R. and Weaver, S. 'Letter to the editor', *Statistics in Medicine*, **15**, 1227–1229 (1996).

