

A Note on Transition Models for Binary 2×2 Cross-Over Data



Kurosawa T¹, Shimokawa A² and Miyaoka E²

¹Department of Applied Mathematics, Tokyo University of Science, Japan

²Department of Mathematics, Tokyo University of Science, Japan

Submission: September 22, 2017; **Published:** October 27, 2017

***Corresponding author:** Takuma Kurosawa, Department of Applied Mathematics, Graduate School of Science, Tokyo University of Science, Japan, Tel: +8190-5329-6165; Email: t.kurosawa1106@gmail.com

Abstract

This paper presents a transition model within the framework of a generalized linear model for binary cross-over data. As a simple example, an analysis of binary 2×2 cross-over data on cerebrovascular deficiency is used to illustrate the model. Several simulation studies are carried out to assess the asymptotic properties of the estimators of the parameters included in the model, with finite sample size, and to compare common methods with respect to the hypothesis test corresponding to the treatment effect. As a main result, we show that the estimators' asymptotic properties work with a moderate sample size, and the transition model takes the highest value with respect to the empirical power among the compared methods.

Keywords: Binary data analysis; Cross-over trial design; Generalized linear model; Partial likelihood

Introduction

The parallel-groups trial design has been studied for decades and is widely used, particularly in medicine and the health sciences. In this design, each unit is randomized to receive a treatment. On the other hand, in a cross-over trial design, each unit receives a sequence of treatments. A cross-over trial design is distinguished from a parallel-groups trial design, although the aim is typically to compare the effects of individual treatments. There are many possible sets of the sequences in a cross-over trial design. The simplest one, called 2×2 design, has two treatments and two periods. In such a design, half the units in the trial receive treatment A and then, after an appropriate period, receive treatment B. The remaining units receive treatment B first and then switch to treatment A. This feature has advantages and disadvantages. As shown in the above 2×2 design example, all units result in a direct comparison of the treatments, because each unit takes some treatments and has the same sample-specific effects. Nevertheless, the comparisons are affected by the sequence of treatments, even if all treatments are identical. Therefore, the main aim of the cross-over trial is to remove the factors related to the differences between the samples from the comparisons.

Various cross-over trial designs have been proposed to remove such factors. For normally distributed 2×2 cross-over data, using two-sample t-tests has been suggested by Chassan [1], and Hills & Armitage [2]. This technique is also applied to linear

models for cross-over data. Jones and Kenward [3] give a good summary of these models. For non-normally distributed 2×2 cross-over data, the t-test can be replaced by nonparametric methods such as the Wilcoxon signed rank test or the Mann-Whitney U-test. These methods were described by Koch [4], Tudor & Koch [5], and Stokes et al. [6]. In the non-normal case, especially for binary cross-over data, Gart [7] applied the Mainland-Gart test to a logistic regression, and Altham [8] derived the same model from a Bayesian point of view. More recently, the test has been applied in various other fields [9,10].

However, these methods use procedures that assume a covariance structure, for example, in generalized estimating equations or generalized linear mixed models [11,12]. An exception to this is the generalized linear transition model used to analyze 2×2 binary cross-over data, as introduced by Miyaoka et al. [13]. In this study, we extend the method based on binary p-ordered transition models within the framework of a generalized linear model and compare it to existing methods used to analyze binary 2×2 cross-over data. Our method need not assume any covariance structures and can naturally incorporate past observations into the model.

Methods

Transition Model for Binary Observations

Let $\{Y_{n,t_n}\}$ be a binary time series, where $n=1, \dots, N$ and $t_n = -p+1, \dots, T_n$ with arbitrary positive integer p . In addition,

let $\{Z_{n,t_n}\}$ be a covariate vector series that can include past observations. For sample n , if the primary event is observed at period t_n then $Y_{n,t_n} = 1$, otherwise, $Y_{n,t_n} = 0$. In keeping with standard practice, we refer to observations 1 and 0 as a success and a failure, respectively. For arbitrary two periods $t_n^1 < t_n^2$, let Y_{n,t_n^1,t_n^2} be $(Y_{n,t_n^1}, \dots, Y_{n,t_n^2})$. Using the simple decomposition rule, the joint probability of $Y_{n,-p+1:t_n}$ is as follows:

$$\Pr(Y_{n,-p+1:t_n}) = \Pr(Y_{n,-p+1:0}) \times \prod_{t_n=1}^{T_n} \Pr(Y_{n,t_n} | Y_{n,-p+1:t_n-1}), \quad (1)$$

Under the assumption that the binary time series has the p -ordered Markov property, the joint probability can be rewritten as follows:

$$\Pr(Y_{n,-p+1:t_n}) = \Pr(Y_{n,-p+1:0}) \times \prod_{t_n=1}^{T_n} \Pr(Y_{n,t_n} | Y_{n,t_n-p:t_n-1}), \quad (2)$$

Let $\Pr(Y_{n,t_n} | Y_{n,t_n-p:t_n-1})$ be $\pi_{n,t_n}^{y_{n,t_n}} (1 - \pi_{n,t_n})^{1-y_{n,t_n}}$, which denotes a success or failure probability, given p observations that were observed in the previous periods.

We further suppose that

$$\pi_{n,t_n} = h_0^{-1}(z'_{n,t_n} \beta), \quad (3)$$

Where h_0^{-1} is an appropriate inverse link function z'_{n,t_n} and represents a transposition of z_{n,t_n} . If the joint probability $\Pr(Y_{n,-p+1:t_n})$ in equation (2) is trivial or easy to find, we can use a full-likelihood estimation. However, in general, the joint probability is difficult to find. Thus, we eliminate the first term on the left-hand side of equation (2), and carry out the estimation based on a partial likelihood. The partial log-likelihood function is expressed as

$$\log PL(\beta) = \sum_{n=1}^N \sum_{t_n=1}^{T_n} \log \left[\left\{ h_0^{-1}(z'_{n,t_n} \beta) \right\}^{y_{n,t_n}} \left\{ 1 - h_0^{-1}(z'_{n,t_n} \beta) \right\}^{1-y_{n,t_n}} \right], \quad (4)$$

Then, the partial score function is given as

$$S_N(\beta) = \nabla \log PL(\beta) = \sum_{n=1}^N \sum_{t_n=1}^{T_n} z_{n,t_n} D_{n,t_n}(\beta) (y_{n,t_n} - \pi_{n,t_n}), \quad (5)$$

Where ∇ denotes a gradient, $D_{n,t_n}(\beta) = \frac{\partial}{\partial \eta} \text{logit}\{h_0^{-1}(\eta)\}$ and $\eta = z'_{n,t_n} \beta$. The corresponding conditional information matrix is

$$G_N(\beta) = \sum_{n=1}^N \sum_{t_n=1}^{T_n} z_{n,t_n} D_{n,t_n}(\beta) \dot{O}_{n,t_n}^{-1} D'_{n,t_n}(\beta) z'_{n,t_n}, \quad (6)$$

Where $\dot{O}_{n,t_n}^{-1} = \text{Var}(Y_{n,t_n} | Y_{n,t_n-p:t_n-1})$ and $D'_{n,t_n}(\beta)$ is the transposed matrix of $D_{n,t_n}(\beta)$. Under normal regularity conditions, the maximum partial likelihood estimator $\hat{\beta}$ is consistent and has asymptotic normality, with $\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, G^{-1}(\beta))$, where $G(\beta) = \lim_{N \rightarrow \infty} \frac{1}{N} G_N(\beta)$. The partial-likelihood proposed by Cox [14], for conditional inference. Note that a more formal definition and theoretical validation for partial-likelihood was given by Wong [15] and Slud [16]. The partial-likelihood and maximum partial likelihood estimator for discrete-valued data

are described in detail in Fokianos & Kedem [17,18]. The small sample behavior of this estimator is discussed in Kurosawa et al. [19].

Transition Model for Binary 2×2 cross-over data

For 2×2 binary cross-over data, let $T_n = 2$ and $p = 1$, equation (2) simplifies to

$$\Pr(Y_{n,0:2}) = \Pr(Y_{n,0}) \times \prod_{t_n=1}^2 \Pr(Y_{n,t_n} | Y_{n,t_n-1}), \quad (7)$$

To calculate partial likelihood, we eliminate $\Pr(Y_{n,0})$ from equation (7) and assume that $Y_{n,1}$ given $Y_{n,0}$ follows a Bernoulli distribution with parameter $\pi_{n,1} = h_1^{-1}(z'_{n,1} \beta)$. For simplicity, let h_0 and h_1 are taken to be logit link functions. Then, the corresponding log-likelihood function is given as

$$l(\beta) = \sum_{n=1}^N \sum_{t_n=1}^2 \log \left[\left\{ h^{-1}(z'_{n,t_n} \beta) \right\}^{y_{n,t_n}} \left\{ 1 - h^{-1}(z'_{n,t_n} \beta) \right\}^{1-y_{n,t_n}} \right] \quad (8)$$

Where h^{-1} implies an inverse logit link function. Although there are various kinds of the covariate vector series $\{Z_{n,t_n}\}$ we use the following parameterization as a simple example:

$$z'_{n,t_n} \beta = \beta_0 + \beta_1 z_{n,t_n,1} + \beta_2 z_{n,t_n,2} + \beta_3 z_{n,t_n,3}, \quad (9)$$

Table 1: 2×2 trial on cerebrovascular deficiency. Observations 0 and 1 correspond to abnormal and normal electrocardiogram readings, respectively.

Group	(0, 0)	(0, 1)	(1, 0)	(1, 1)	Total
1	12	2	7	29	50
2	13	6	5	26	50
Total	25	8	12	55	100

Where $z_{n,t_n,1}$ is 1 if sample n received treatment A in period t_n and 0 otherwise, and $(z_{n,t_n,2}, z_{n,t_n,3})$ is $(y_{n,t_n-1}, 1 - y_{n,t_n-1})$ if $t_n \geq 2$ and otherwise both are 0. In this situation, parameter β_1 denotes the treatment effect and the principle interest is to test the hypothesis $H_0: \beta_1 = 0$. The parameters β_2 and β_3 represent a period effect and a test $H_0: \beta_2 = \beta_3$ indicates the absence of a period effect. Parameter β_0 is the intercept. To illustrate the use of the transition model for 2×2 binary cross-over data, we apply the data published by Jones & Kenward [3] to the model described above. Their original data were divided into two centers, however, in this example, we use data that were merged without considering differences in centers. The samples are randomized into two groups, namely group 1 and group 2. In group 1, each sample received treatment A at period 1 and treatment B at period 2. In group 2, each sample first received treatment B, and then treatment A. In Table 1, (i, j) denotes an ordered pair of two binary observations, where i and j represent the observation at period 1 and period 2, respectively.

The results are presented in Table 2. Focusing on β_1 in the table, the effect of treatment A is likely to be positive. However, the Wald test statistic of 1.440 for the hypothesis $\beta_1 = 0$

corresponding to a p-value of 0.075, leads us to conclude that there is insufficient evidence of a treatment effect. On the other hand, the likelihood ratio test statistic is 33.92 for the hypothesis $\beta_2 = \beta_3 = 0$, and the corresponding p-value is smaller than 0.001.

Table 2: Estimated $\hat{\beta}_i$ and standard errors in cerebrovascular deficiency example.

Parameter	Estimate	Estimated SE
β_0	0.48	0.262
β_1	0.477	0.331
β_2	0.84	0.386
β_3	-1.908	0.465

Results

Simulation results

We show two simulation results in this section. The first assesses the asymptotic properties of $\hat{\beta}$ for a finite sample size, as described at transition model for binary observations, while the second compares the tendency to reject the principle interest hypothesis, which implies that the effects are different between

treatments A and B. For all the simulations, we generated data 10,000 times for $N = 40, 60, 80, 100$ and 200 samples and the N samples were allocated randomly to two equally sized groups in each simulated data set.

Simulation study I

Here, we assessed the asymptotic properties of $\hat{\beta}$ for a finite sample size. The simulated data were generated from a Bernoulli distribution with parameter π_{n,t_n} calculated from equation (9). The parameter values $\hat{\beta} = (0.478, 0.477, 0.840, -1.908)$ were chosen from the previous example. The mean squared errors of $\hat{\beta}_i$ are summarized in Table 3. As shown in Table 3, $\hat{\beta}_0$ and $\hat{\beta}_1$ are well approximated with a small sample size. On the other hand, a moderate to high sample size is required for $\hat{\beta}_2$ and $\hat{\beta}_3$ to converge to their true values. These differences depend on whether the covariate is time dependent. Figure 1 displays histograms and normal quantile-quantile plots of the simulated $\hat{\beta}_i$ for each sample size and indicates that the normal approximation fits well, as does the mean squared error. We also show histograms and normal quantile-quantile plots for $\hat{\beta}_0$, $\hat{\beta}_2$ and $\hat{\beta}_3$ and confirm convergence, with the same tendency as consistency.

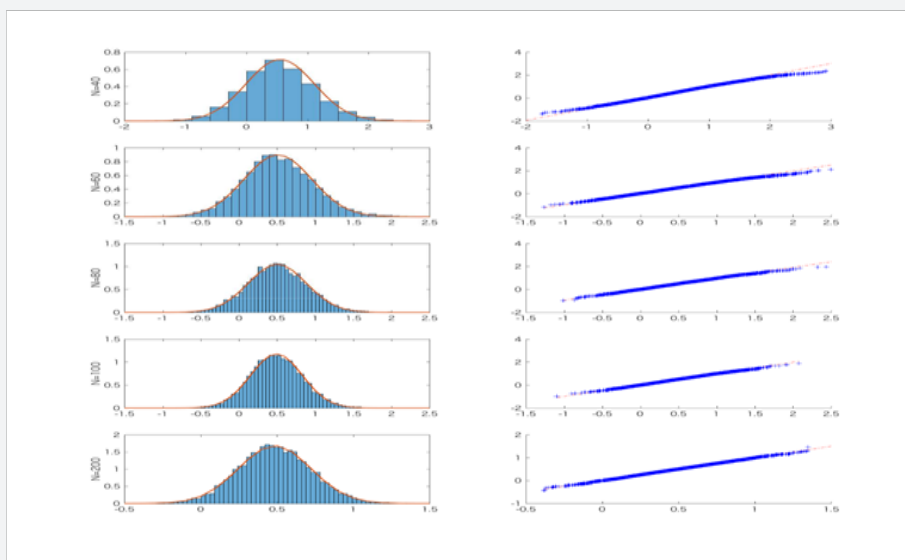


Figure 1: Histograms(left column) and normal quantile-quantile plots(right column) of simulated $\hat{\beta}_i$ in simulation study I. For histograms, the horizontal axis shows value of simulated $\hat{\beta}_i$ and the vertical axis shows its estimated density.

Table 3: Mean squared error of $\hat{\beta}_i$ in simulation study I.

N	MSE of $\hat{\beta}_0$	MSE of $\hat{\beta}_1$	MSE of $\hat{\beta}_2$	MSE of $\hat{\beta}_3$
40	0.195	2.427	59	356.3
60	0.127	0.216	2.355	57.45
80	0.092	0.158	2.249	8.639
100	0.073	0.122	0.17	4.415
200	0.035	0.056	0.078	0.117

Simulation study II

In this simulation, we compare the following four methods: the conditional Likelihood approach (CL), generalized linear mixed model (GLMM), generalized estimating equation (GEE), and transition model (TM). The simulated data were generated from a Bernoulli distribution with parameter p_{n,t_n} which was satisfied following equation:

$$\text{logit}(p_{n,t_n}) = \xi_{t_n} + \phi_l + \psi_{l'} + \delta_n + \varepsilon_{n,t_n}, \quad (10)$$

Where ξ_{t_n} for $t_n = 1, 2$ is an effect in period t_n , ϕ_l for $l = A, B$ is the direct effect of treatment l , ψ_l for $l = A, B$ is the carryover effect of treatment l , δ_n is a subject-specific effect, and ε_{n,t_n} is a random error. We assume that the TM is defined as equations (7)-(9), with the other models defined as follows. For the CL and GLMM,

$$\text{logit}\left(\Pr(Y_{n,t_n}=1|b_n)\right)=\gamma_0+\gamma_1x_{n,t_n,1}+\gamma_2x_{n,t_n,2}+\gamma_3x_{n,t_n,3}+b_n, \quad (11)$$

and for the GEE,

$$\text{logit}\left(\Pr(Y_{n,t_n}=1)\right)=\gamma_0^M+\gamma_1^Mx_{n,t_n,1}+\gamma_2^Mx_{n,t_n,2}+\gamma_3^Mx_{n,t_n,3} \quad (12)$$

Where $\tilde{\alpha}_1, \gamma_2, \gamma_3$ and b_n denote the treatment effect, period effect, sequence effect, and subject-specific effect, respectively, and the superscript M denotes the marginal model parameter. In equations (11) and (12), $x_{n,t_n,1}$ is 1 if sample n received treatment A in period t_n and is 0 otherwise, $x_{n,t_n,2} = t_n$ and $x_{n,t_n,3}$ is 1 if sample n received treatment A in period 1, and is 2 otherwise. Because the CL and GLMM estimate the subject-specific effect, whereas the GEE and TM estimate the population average, the above four models differ in terms of their interpretation. In addition, because some of the parameters in the model are not common, there is no point in directly comparing those values. Therefore, in this simulation, we compare the four methods by testing the hypothesis on the parameter representing the treatment effect. We set the

null hypothesis to $H_0: \beta_1 = 0$ $H_0: \gamma_1^M = 0$ for TM, for GEE, or $H_0: \gamma_1 = 0$ or the other methods and conduct the Wald test with size $\alpha = 0.05$. We further assume that parameter vector takes $\theta = (\xi_1, \xi_2, \phi_A, \phi_B, \psi_A, \psi_B) = (0.10, 0.05, 0.70, 0.00, 0.05, 0.03)$ or $(0.10, 0.05, 0.70, 0.70, 0.05, 0.03)$ and ε_{n,t_n} that the values of b_n and ε_{n,t_n} are generated independently from a standard normal distribution.

Table 4: Empirical power of the Wald test with size over 10,000 simulations under the alternative ($\beta_1 \neq 0$ for TM, $\tilde{\alpha}_1^M \neq 0$ for GEE, $\tilde{\alpha}_1 \neq 0$ for the other methods) in simulation study II.

N	CL	GLMM	GEE	TM
40	0.194	0.228	0.226	0.3
60	0.396	0.393	0.379	0.42
80	0.517	0.491	0.504	0.528
100	0.62	0.592	0.616	0.624
200	0.843	0.82	0.864	0.847

Table 5: Empirical type I error rate of the Wald test with size over 10,000 simulations under the null ($\beta_1 = 0$ for TM, $\tilde{\alpha}_1^M = 0$ for GEE, $\tilde{\alpha}_1 = 0$ for the other methods) in simulation study II.

N	CL	GLMM	GEE	TM
40	0.031	0.033	0.035	0.055
60	0.042	0.038	0.047	0.049
80	0.048	0.038	0.043	0.05
100	0.048	0.034	0.048	0.05
200	0.043	0.038	0.045	0.047

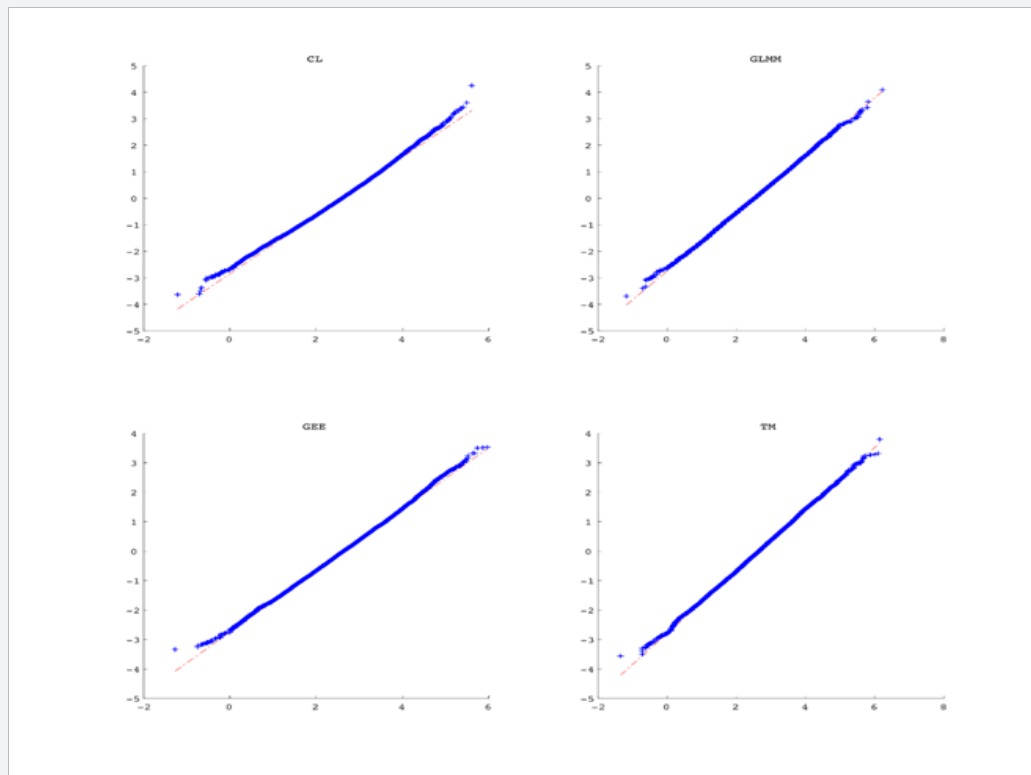


Figure 2: Normal quantile-quantile plots of the Wald test statistics when the treatment effect is different between A and B when the sample size is 40 and 200 in simulation study II.

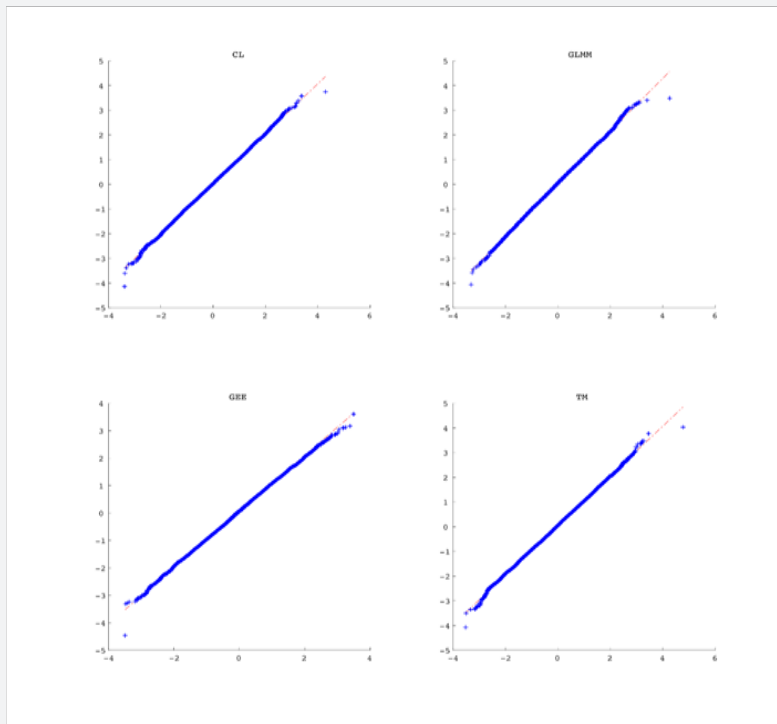


Figure 3: Normal quantile-quantile plots of the Wald test statistics when the treatment effect is the same between A and B when the sample size is $n = 10, 20, 50, 100, 200, 500, 1000$ and $\theta = (0.10, 0.05, 0.70, 0.00, 0.05, 0.03)$ in simulation study II.

The former is the case in which the treatment effect is different between A and B, while the later is the case in which the treatment effect is the same for treatment A and B. Table 4 summarizes the empirical power of the alternatives using the Wald test over the 10,000 simulations. As presented in Table 4, the TM has the highest value, although only slightly. In the case of a small sample size, the CL has the smallest value, while, for a sufficiently large sample size, all the methods take similar values. Table 5 shows the empirical type-I error rate of the Wald test over the 10,000 simulations. This table shows that the GLMM has most conservative results, this is because GLMM overestimates the standard error compared to other methods. For a small sample size, the TM inflates the type-I error rate, although it never exceed the nominal level for a moderate sample size. As with the case of the empirical power, the methods take a similar empirical type-I error rate in the case of a sufficiently large sample size. Normal quantile-quantile plots of the Wald statistic for each case are summarized in Figures 2 & 3. These plots show that the Wald statistic for each method is well approximated.

Conclusion

In this study, we constructed a transition model for binary 2×2 cross-over data based on a generalized linear model and conducted several simulation studies. The results of the simulation studies are as follows.

First, the estimators of the parameters included in the transition model have consistency and asymptotic normality with a moderate to high sample size. Thus, we were able to

conduct statistical hypothesis tests, such as the Wald test and the likelihood ratio test, using the transition model with finite sample size as with several methods used to analyze binary 2×2 cross-over data in general. Secondly, the transition model takes the highest value with respect to the empirical power, and except for the case of a small sample size, the empirical type-I error rate is lower than the nominal level as with the other methods.

This implies that other methods overestimate the variance of the estimated parameter. We conclude that, as with other methods used to analyze 2×2 cross-over data, the transition model can perform hypothesis tests on matters of interest and, unlike the generalized estimating equations and generalized linear mixed model, there is no need to assume any variance covariance structures. Therefore, the transition model can be described simple model and does not overestimate the variance of estimated parameters. In addition, it is possible to include covariates associated with past responses into the model and to deal naturally with more complicated cross-over designs that have more than two periods.

References

1. Chassan JB (1964) On the Analysis of Simple Cross-overs with Unequal Numbers of Replicates. *Biometrics* 20(1): 206-208.
2. Hills M, Armitage P (1979) The Two-Period Cross-Over Clinical Trial. *British Journal of Clinical Pharmacology* 8(1): 7-20.
3. Jones B, Kenward GM (2015) Design and Analysis of Cross-Over Trials(3rd edn), Chapman and Hall: London, UK.
4. Koch GG (1972) The Use of Non-Parametric Methods in the Statistical

- Analysis of the Two-Period Changeover Design. *Biometrics* 28(2): 577-584.
5. Tudor GE, Koch GG (1994) Review of Nonparametric Methods for the Analysis of Crossover Studies. *Statistical Methods in Medical Research* 3(4): 345-381.
 6. Stoke MS, Davis CS, Koch GG (2012) *Categorical Data Analysis Using the SAS System* (3rd edn), SAS Institute Inc.: Cary, NC, USA.
 7. Gart JJ (1969) An Exact Test for Comparing Matched Proportions in Crossover Designs. *Biometrika* 56(1): 75-80.
 8. Altham EMP (1971) The Analysis of Matched Proportions. *Biometrika* 58 (3): 561-576.
 9. Diggle PJ, Heagerty P, Liang KY, Zeger SL (2002) *Analysis of Longitudinal Data*. (2nd edn), Oxford University Press: Oxford, UK.
 10. Molenberghs G, Verbeke G (2005) *Models for Discrete Longitudinal Data*, New York: Springer-Verlag, USA.
 11. Hu BF, Goldberg J, Hedeker D, Flay RB, Pentz AM (1998) Comparison of Population-Averaged and sample-Specific Approaches for Analyzing Repeated Binary Outcomes. *American Journal of Epidemiology* 147(7): 694-703.
 12. Lee Y, Nelder AJ (2004) Conditional and Marginal Models: Another View. *Statistical Science* 19(2): 219-238.
 13. Miyaoka E, Kajikawa M, Smith B (2000) Analysis of Binary Cross-Over Data with Generalized Linear Transition Models. *Journal of Japan Statistical Society* 30(2000): 17-25.
 14. Cox DR (1975) Partial Likelihood. *Biometrika* 62 (2): 269-276.
 15. Wong WH (1986) Theory of Partial Likelihood. *The Annals of Statistics* 14(1): 88-123.
 16. Slud E (1992) Partial Likelihood for Continuous-time Stochastic Processes. *Scandinavian Journal of Statistics* 19: 97-109.
 17. Fokianos K, Kedem B (1998) Prediction and classification of Non-Stationary Categorical time series. *Journal of Multivariate Analysis* 67(2): 277-296.
 18. Kedem B, Fokianos K (2002) *Regression Models for Time Series Analysis*, New York: Wiley.
 19. Kurosawa T, Shimokawa A., and Miyaoka E. (2016). A Note on Using the Maximum Partial Likelihood Estimator of Transition Model for Binary Time Series. *Advances and Applications in Statistics* 48(1): 69-90.



This work is licensed under Creative Commons Attribution 4.0 License
DOI: [10.19080/BBOAJ.2017.03.555622](https://doi.org/10.19080/BBOAJ.2017.03.555622)

Your next submission with Juniper Publishers

will reach you the below assets

- Quality Editorial service
- Swift Peer Review
- Reprints availability
- E-prints Service
- Manuscript Podcast for convenient understanding
- Global attainment for your research
- Manuscript accessibility in different formats (Pdf, E-pub, Full Text, Audio)
- Unceasing customer service

Track the below URL for one-step submission

<https://juniperpublishers.com/online-submission.php>