SUT Journal of Mathematics
Vol. 57, No. 2 (2021), 147–165

Non-inferiority marginal symmetry model and its decomposition for ordinal square contingency tables

Mana Aizawa, Shuji Ando, Kouji Tahata and Sadao Tomizawa

(Received April 14, 2021)

Abstract. The marginal homogeneity (MH) model is well-known for analyzing ordinal square contingency tables. This study proposes a non-inferiority marginal symmetry (NiMS) model, which has a different marginal symmetry structure than the MH model. In the NiMS model, the probability of an observation falling in row category i or below and column category i or above is equal to the probability of an observation falling in row category i or above and column category i or below. Additionally, two kinds of extended NiMS models are proposed. These extended NiMS models constantly hold when the NiMS model holds. However, the converse is not necessarily true. This study examines what a model should be necessary, in addition to the extended NiMS model, to satisfy the NiMS model.

AMS 2010 *Mathematics Subject Classification.* 62H17

Key words and phrases. Asymmetry, clinical study, ordered category, generalized marginal symmetry, marginal homogeneity, relationships among models

§1. Introduction

Typically, the statistical independence does not hold between row and column variables in an $r \times r$ ordinal square contingency table with the same row and column classifications. Because observations tend to concentrate on the main diagonal cells, the symmetry or asymmetry between the row and column variables tends to be of more interest than the independence. Many models have been proposed to examine the symmetry or asymmetry. Examples of models focusing on the symmetry include the symmetry (S) model (Bowker, 1948), the quasi-symmetry model (Caussinus, 1965), and the marginal homogeneity (MH) model (Stuart, 1955). On the other hand, the conditional symmetry (CS) model (McCullagh, 1978), the extended marginal homogeneity (EMH) model (Tomizawa, 1984), and the diagonals-parameter symmetry model (Goodman, 1979) consider the asymmetry.

Let p_{ij} denote the probability that an observation will fall in the ith row and jth column of the table ($i = 1, \ldots, r; j = 1, \ldots, r$). Let X and Y denote the row and column variables, respectively. The S model is defined by

$$p_{ij} = p_{ji} \quad (i = 1, \ldots, r; j = 1, \ldots, r).$$

The S model indicates that the symmetric structure of the cell probabilities $\{p_{ij}\}$ with respect to the main diagonal of the table.

The CS model is an extension of the S model with an asymmetric structure and is defined by

$$p_{ij} = \Delta p_{ji} \quad (1 \leq i < j \leq r).$$

The CS model indicates that the probability that X is i and Y is j, for $i < j$, is Δ times higher than the probability that X is j and Y is i. When $\Delta = 1$, the CS model is equivalent to the S model. The CS model constantly holds when the S model holds. However, the converse is not necessarily true. Read (1977) considered the global symmetry (GS) model defined by

$$\sum_{i<j}\sum p_{ij} = \sum_{i>j}\sum p_{ij},$$

and gave the decomposition that the S model holds if and only if both the CS and GS models hold.

The MH model is defined by

$$p_{i\cdot} = p_{\cdot i} \quad (i = 1, \ldots, r),$$

where $p_{i\cdot} = \sum_{k=1}^{r} p_{ik}$ and $p_{\cdot i} = \sum_{k=1}^{r} p_{ki}$. The MH model indicates that the marginal distribution of the row variable is identical to the marginal distribution of the column variable. When $r = 2$, the MH model is equivalent to the S model. For ordinal square contingency tables, the MH model can also be expressed as

$$G_{1(i)} = G_{2(i)} \quad (i = 1, \ldots, r-1),$$

where

$$G_{1(i)} = \Pr(X \leq i, Y \geq i+1), \quad G_{2(i)} = \Pr(X \geq i+1, Y \leq i).$$

This indicates that the probability that X is i or below and Y is $i+1$ or above equals to the probability that X is $i+1$ or above and Y is i or below for $i = 1, \ldots, r-1$.

Many authors have proposed extensions of the MH model. Examples include Agresti (1984, Sec 11.1), Iki, Tahata, and Tomizawa (2010), Yamamoto, Shinoda, and Tomizawa (2011), and Saigusa, Maruyama, Tahata,

and Tomizawa (2018). For example, Tomizawa (1984) defined the EMH model as

$$G_{1(i)} = \Delta G_{2(i)} \quad (i = 1, \ldots, r-1).$$

The EMH model indicates that the probability that X is i or below and Y is $i+1$ or above is Δ times higher than the probability that X is $i+1$ or above and Y is i or below for $i = 1, \ldots, r-1$. The EMH model with $\Delta = 1$ is equivalent to the MH model. The EMH model constantly holds when the MH model holds. However, the converse is not necessarily true. Tomizawa (1991) considered the marginal mean equality (ME) model defined by $E(X) = E(Y)$, where $E(X) = \sum_{i=1}^{r} ip_{i\cdot}$ and $E(Y) = \sum_{i=1}^{r} ip_{\cdot i}$, and gave the decomposition that the MH model holds if and only if both the EMH and ME models hold.

Tomizawa (1995) proposed the generalized marginal homogeneity (GMH) model, which is defined as

$$G_{1(i)} = \Delta \Phi^{i-1} \, G_{2(i)} \quad (i = 1, \ldots, r-1).$$

The GMH models with $\Phi = 1$ and $\Delta = \Phi = 1$ are equivalent to the EMH and MH models, respectively.

The MH model and its extensions are often used to analyze agreement/disagreement between ratings of paired evaluations (e.g., assessment of pre-post treatment for drug efficacy/safety or evaluation of right and left eye vision grade). The EMH and GMH models can analyze which paired rating is better.

In this study, we focus on the situation where the treatment is effective even if the rating is unchanged within a subject. For example, we consider the quality of life (QoL) data taken from Clark (2019). These data were collected in a pressure ulcer prevention study on spinal cord injuries (SCIs). Table 1 is the data of shifts in the RAND 36-Item Health Survey (SF-36) general health subscale score taken 12 months from the baseline in patients with SCIs who received the intervention.

As Carlson et al. (2019) mentioned, medically serious pressure injuries (MSPrIs) are common complication of SCIs. They have devastating consequences on health and well-being and are extremely expensive to treat. Therefore, an assessment of the maintenance or improvement of patient's QoL is meaningful to evaluate the effectiveness of an intervention.

For such data (Table 1), a comparison of the post- and pre-treatment status is of interest. The EMH and GMH models can consider whether the post-treatment status is better than the pre-treatment one. However, they do not consider whether the post-treatment status is the same as or better than the pre-treatment one. To tackle this issue, this study proposes new models. Moreover, this study gives decompositions of new models.

The remainder of this paper is organized as follows. Section 2 proposes new models. Section 3 gives their decompositions. Section 4 describes the goodness-of-fit test. Section 5 shows examples by applying the new models to real data from clinical studies. Section 6 and Section 7 present the discussion and concluding remarks, respectively.

§2. Non-inferiority marginal symmetry model

In this section, we propose three kinds of new models, which can easily interpret situations where the treatment is meaningful even if the rating is unchanged/maintained within a subject.

First, we propose the non-inferiority marginal symmetry (NiMS) model, which is defined by

$$W_{1(i)} = W_{2(i)} \quad (i = 1, \ldots, r),$$

where

$$W_{1(i)} = \Pr(X \leq i, Y \geq i), \quad W_{2(i)} = \Pr(X \geq i, Y \leq i).$$

The NiMS model indicates that the probability that X is i or below and Y is i or above equals to the probability that X is i or above and Y is i or below for $i = 1, \ldots, r$. When $r = 3$, the NiMS model is equivalent to the S model. The NiMS model should be used when $r \geq 3$.

Second, similar to the relationship between the MH and EMH models, we propose the extended NiMS (ENiMS) model, which is defined by

$$W_{1(i)} = \Delta W_{2(i)} \quad (i = 1, \ldots, r).$$

The ENiMS model with $\Delta = 1$ is equivalent to the NiMS model. The ENiMS model indicates that the ratio of $W_{1(i)}$ and $W_{2(i)}$ is constant (i.e., Δ) for all $i = 1, \ldots, r$.

Third, similar to the relationship between the MH and GMH models, we propose the generalized NiMS (GNiMS) model, which is defined by

$$W_{1(i)} = \Delta \Phi^{i-1} W_{2(i)} \quad (i = 1, \ldots, r).$$

The GNiMS models with $\Phi = 1$ and $\Delta = \Phi = 1$ are equivalent to the ENiMS and NiMS models, respectively. The GNiMS model indicates that the ratio of $W_{1(i+1)}/W_{2(i+1)}$ and $W_{1(i)}/W_{2(i)}$ is constant (i.e., Φ) for $i = 1, \ldots, r-1$.

Consider an ordinal square contingency table constructed by the assessment of pre- and post-treatment for drug efficacy/safety. The NiMS model indicates that compared to the pre-treatment status, the probability that the

post-treatment status is maintained or improved equals the probability that the post-treatment status is deteriorated or maintained. On the other hand, the ENiMS model indicates that compared to the pre-treatment status, the probability that the post-treatment status is maintained or improved is Δ times higher than the probability that the post-treatment status is deteriorated or maintained. Thus, the ENiMS model assumes that the treatment effect is always constant and independent of the pre-treatment status. Additionally, the GNiMS model indicates that compared to the pre-treatment status, the probability that the post-treatment status is maintained or improved is $\Delta\Phi^{i-1}$ times higher than the probability that the post-treatment status is deteriorated or maintained. Thus, the GNiMS model assumes that the treatment effect depends on the pre-treatment status.

§3. Decompositions of the model

We assign the score given by the real-valued function of two variables g to the pair of (x, y). Let $Z = g(X, Y)$. Consider the model where $E(Z) = 0$. We shall refer to this model as the zero mean based on g (ZM[g]) model.

Define the function g_{GS} as

$$g_{GS}(x, y) = \begin{cases} 1 & (x < y), \\ -1 & (x > y), \\ 0 & (x = y). \end{cases}$$

Namely, the function g_{GS} gives (1) a score of 1 for right upper triangle cells, (2) a score of -1 for left lower triangle cells, and (3) a score of 0 for main diagonal cells. Then, the ZM[g_{GS}] model can be expressed as

$$\sum_{i<j}\sum p_{ij} - \sum_{i>j}\sum p_{ij} = 0.$$

Namely, the ZM[g_{GS}] model is equivalent to the GS model. Additionally, define the function g_{ME} by

$$g_{ME}(x, y) = y - x.$$

The score of $g_{ME}(x, y)$ is related to the distance from main diagonal of the square table. Since the ZM[g_{ME}] model can be expressed as $E(Y - X) = 0$, it is equivalent to the ME model. That is, $E(X) = E(Y)$.

Generally, the function g is related to the difference $y - x$. For the specified

constant d, we consider

$$g_d(x,y) = \begin{cases} y - x + d & (x < y), \\ -(x - y + d) & (x > y), \\ 0 & (x = y). \end{cases}$$

We note that $g_0 = g_{ME}$ and d may be useful to emphasize the difference between main diagonal cells and non-main diagonal cells. Here, we focus on the function $g_1(x,y)$. The ZM[g_1] model can be expressed as

$$\sum_{t=1}^{r} W_{1(t)} - \sum_{t=1}^{r} W_{2(t)} = 0.$$

We obtain the following theorem.

Theorem 1. *The NiMS model holds if and only if both the ENiMS and ZM[g_1] models hold.*

Proof. If the NiMS model holds, it is obvious that both the ENiMS and ZM[g_1] models hold. Assume that both the ENiMS and ZM[g_1] models hold. From the ENiMS model,

$$\sum_{t=1}^{r} W_{1(t)} - \sum_{t=1}^{r} W_{2(t)} = \Delta \sum_{t=1}^{r} W_{2(t)} - \sum_{t=1}^{r} W_{2(t)} = (\Delta - 1) \sum_{t=1}^{r} W_{2(t)}.$$

Since the ZM[g_1] model holds, we obtain $\Delta = 1$. Namely, the NiMS model holds. The proof is complete. □

Now consider the following structure of the concordance (C) and the discordance (D), which are given as

$$C = \sum \sum_{s<t} W_{1(s)} W_{2(t)}, \quad D = \sum \sum_{s<t} W_{2(s)} W_{1(t)}.$$

We shall refer to the following model as the concordance-discordance equality (CDE) model;

$$C = D.$$

Then, we obtain the following theorem.

Theorem 2. *The ENiMS model holds if and only if both the GNiMS and CDE models hold.*

Proof. Assume that the ENiMS model holds. It is obvious that the GNiMS model holds. From the ENiMS model, we have

$$C = \sum\sum_{s<t} W_{1(s)}W_{2(t)} = \Delta \sum\sum_{s<t} W_{2(s)}W_{2(t)}$$

and

$$D = \sum\sum_{s<t} W_{2(s)}W_{1(t)} = \Delta \sum\sum_{s<t} W_{2(s)}W_{2(t)}.$$

Thus, we also obtain $C = D$. Namely, the CDE model holds.

Conversely, we assume that both the GNiMS and CDE models hold. From the GNiMS model, we have

$$C = \sum\sum_{s<t} W_{1(s)}W_{2(t)} = \Delta \sum\sum_{s<t} \Phi^{s-1}W_{2(s)}W_{2(t)}$$

and

$$D = \sum\sum_{s<t} W_{2(s)}W_{1(t)} = \Delta \sum\sum_{s<t} \Phi^{t-1}W_{2(s)}W_{2(t)}.$$

Since the CDE model holds, we obtain $\Phi = 1$. Namely, the ENiMS model holds. The proof is complete. □

From Theorems 1 and 2, we obtain the following corollary.

Corollary 1. *The NiMS model holds if and only if all the GNiMS, ZM[g_1], and CDE models hold.*

§4. Goodness-of-fit test

Let n_{ij} denote the observed frequency in the (i,j)th cell of the table ($i = 1,\ldots,r$; $j = 1,\ldots,r$). We assume multinomial sampling over the cells of the table. Thus, the observed frequencies $\{n_{ij}\}$ have a multinomial distribution with the cell probabilities $\{p_{ij}\}$ as parameters. The maximum likelihood estimates (MLEs) of the expected frequencies under the NiMS, ENiMS, and GNiMS models are obtained, for example, using the Newton-Raphson method in the log-likelihood equation. See Appendix for details. The likelihood ratio chi-squared statistics for testing goodness-of-fit of the model M is given by

$$G^2(M) = 2\sum_{i=1}^{r}\sum_{j=1}^{r} n_{ij}\log\left(\frac{n_{ij}}{\hat{m}_{ij}}\right),$$

where $\hat{m}_{ij}$ is the MLE of the expected frequency m_{ij} under the model M. The number of degrees of freedom (df) for the NiMS model is r where $r \geq 3$.

The numbers of df for the ENiMS and GNiMS models are $r-1$ and $r-2$, respectively. Note that the numbers of df for the MH, EMH, and GMH are $r-1$, $r-2$ and $r-3$, respectively.

Assume that the models M_1 and M_2 are nested and the model M_1 has fewer parameters than M_2 (i.e., if the model M_1 holds, then the model M_2 also holds). For testing that the model M_1 holds assuming that the model M_2 holds true, the likelihood ratio statistics is given as

$$G^2(M_1|M_2) = G^2(M_1) - G^2(M_2).$$

Under the null hypothesis, the statistics $G^2(M_1|M_2)$ has an asymptotic chi-squared distribution where the number of df that is equal to the difference between the numbers of df for the models M_1 and M_2. For example, see Agresti (2019, Sec 3.4.4).

The test statistics $G^2(M_1|M_2)$ cannot be used to compare non-nested models. Thus, other statistics are necessary for such a comparison. A well-known example to compare non-nested models is the Akaike Information Criterion (AIC) (Akaike, 1974). The AIC is defined as

$$\text{AIC} = -2(\text{maximum log likelihood}) + 2(\text{number of parameters}).$$

The AIC indicates that the best fitting model is the one with the minimum AIC. For details, see Konishi and Kitagawa (2008, Sec 3.4.4). Since only the difference between AICs is necessary to compare models, the common constant in AIC can be ignored. Thus, the modified AIC (AIC^+) for the model M is defined as

$$\text{AIC}^+(M) = G^2(M) - 2\text{df}.$$

§5. Applications to real data from clinical studies

5.1. Example 1

Table 1 shows data collected in a pressure ulcer prevention study in patients with SCIs (Clark, 2019). This study is a randomized controlled study to evaluate the efficacy of the lifestyle-based intervention designed to reduce incidence of MSPrIs for adult patients with SCIs. The intervention in this study is entitled Pressure Ulcer Prevention Program (PUPP). Table 1 shows the shifts in the SF-36 general health subscale score at 12 months from the baseline in patients with SCIs who received the PUPP intervention. The SF-36 subscale score is a health-related QoL indicator. The SF-36 has eight subscales for health concepts, and one is general health. In Table 1, the row and column variables describe the SF-36 general health subscale score at the baseline and

at 12 months, respectively. The original scale ranges from 0 to 100, where 0 and 100 indicate the maximum disability and no disability, respectively. We divided the SF-36 subscale score into four categories (i.e., (1): $0- \leq 25$, (2): $> 25- \leq 50$, (3): $> 50- \leq 75$, and (4): $> 75- \leq 100$), where (1) is the worst and (4) is the best.

From Table 2, since the NiMS, ENiMS and GNiMS models fit the data in Table 1 well, we compare the goodness-of-fit of these nested models using the likelihood ratio statistics shown in Section 4. The NiMS model is preferable to the ENiMS and GNiMS models, this is because $G^2(\text{NiMS}|\text{ENiMS}) = 0.940$ and $G^2(\text{NiMS}|\text{GNiMS}) = 1.009$.

From Table 2, we see that all models applied to data in Table 1 fit well. Since the MH model that is non-nested with the NiMS model includes in these models, we compare the goodness-of-fit of these models using the AIC^+. The NiMS model, which has the minimum AIC^+ among models applied to data in Table 1, is the best fitting model.

Under the NiMS model, the probability of maintaining or improving the SF-36 general health subscale is the same as the probability of maintaining or deteriorating the SF-36 general health subscale in patients with SCIs who received PUPP intervention. Under the ENiMS model, the MLE of Δ is 1.190. According to the test based on $G^2(\text{NiMS}|\text{ENiMS})$, the hypothesis of $\Delta = 1$ cannot be rejected at the significance level of 0.05. Therefore, the PUPP intervention tends to maintain or improve SF-36 general health subscale score but it is not significant.

5.2. Example 2

Consider the data in Table 3 taken from Schuette et al. (2012). These data were collected in the clinical study entitled performance status and influencing factors during second-line treatment with pemetrexed in patients with stage III/IV non-small cell lung cancer (NSCLC). This study is a prospective, non-interventional phase IV multicenter study to evaluate the changes in physician-rated performance status and patient-rated health related QoL during second-line treatment with pemetrexed in routine clinical practice.

As Schuette et al. (2012) mentioned, the most important purpose of the second-line treatment for patients with NSCLC Stage IIIa/b or IV is palliation. Therefore, the maintenance or improvement of patient's overall health condition measured by the performance status is highly relevant. Since the performance status is an important prognostic factor, the treatment effect may depend on the performance status at pre-treatment (i.e., having a poor performance status at the pre-treatment is associated with a poor prognosis for treatment).

Table 3 shows the data obtained by cross-classifying for the Karnofsky Index (KI) at the baseline and the KI after the second treatment cycle with pemetrexed. The KI is a score to measure performance status for cancer patients. Note that the range of the KI is from 0% to 100%, which is observed in 10% increments and not a continuous value. A KI of 100% means "Normal, No complaints", and a KI of 0% means "Dead". We categorized the KI into four categories similar to Schuette et al. (2012): (1) $\geq 80\%$, (2) 70%, (3) 60% and (4) $\leq 50\%$. (1) is the best condition and (4) is the worst condition.

From Table 4, the GNiMS model fits the data well, where the MLE of Δ is 1.078 and Φ is 1.421. Therefore, under the GNiMS model, the probability that the patient's KI at the baseline is categorized as i or below and his/her KI after the second cycle treatment is categorized as i or above (i.e., the KI unchanged or deteriorated) is estimated to be $1.078 \times 1.421^{i-1}$ times higher than the probability that the patient's KI at the baseline is categorized as i or above and his/her KI after the second cycle treatment is categorized as i or below (i.e., the KI unchanged or improved) for $i = 1, 2, 3, 4$ in Table 3. When $i = 4$, the probability that the KI remains unchanged or deteriorated is 3.093 times higher than the probability that the KI remains unchanged or improved. Thus, the probability that the KI after the second treatment cycle is maintained or deteriorated from the baseline is higher than the probability that the KI after the second treatment cycle is maintained or improved from the baseline in patients with NSCLC Stage IIIa/b or IV who are treated with pemetrexed. In addition, this tendency becomes stronger in patients whose baseline KI is poorer. In other words, the tendency is weak in patients whose baseline KI is well.

Table 4 shows that the ENiMS model fits the data in Table 3 poorly. We also see the CDE model fits poorly, whereas the GNiMS model fits well. Therefore, from Theorem 2, the poor fit of the ENiMS model is due to the lack of structure of the CDE model rather than the GNiMS model. Similarly, from Table 4, we see the NiMS and ZM[g_1] models fit the data in Table 3 poorly. Thus, from Corollary 1, the poor fit of the NiMS model is due to the lack of structure of the CDE and ZM[g_1] models rather than the GNiMS model.

§6. Discussion

Read (1977) noted that the S model holds if and only if both the CS and GS models hold. Tomizawa (1991) gave that the MH model holds if and only if both the EMH and ME models hold. In this paper, we showed that the NiMS model holds if and only if both the ENiMS and ZM[g_1] models hold. As described in Section 3, the GS and ME models are special cases of the ZM[g] model. Namely, the zero mean model based on the function g is useful

to consider the equivalence condition of symmetry (or homogeneity).

Many studies have proposed extensions of the MH model. Therefore, additional expansions of the NiMS model should be considered. For example, Tahata and Tomizawa (2008) proposed the m-additional parameter marginal homogeneity (MH(m)) model for fixed m ($m = 1, 2, \ldots, r-1$), which is defined by

$$G_{1(i)} = \Delta_i^{(m)} \, G_{2(i)} \quad (i = 1, \ldots, r-1),$$

where $\Delta_i^{(m)} = \prod_{k=0}^{m-1} \delta_k^{i^k}$. The MH(1) and MH(2) models are equivalent to the EMH and GMH models, respectively. Similar to the MH(m) model, the m-additional parameter NiMS model should be further evaluated to consider the m-additional parameter.

§7. Concluding remarks

The MH model and its extensions are often used to analyze agreement/disagreement between rating of paired evaluations. The EMH and GMH models can analyze which paired rating is better. However, they do not consider that one of the paired evaluation is the same as or better than the another one. To tackle this issue, we proposed the NiMS, ENiMS, and GNiMS models. These proposed models can easily interpret situations where the paired evaluation is meaningful even if the rating is unchanged/maintained within the pair. When we consider the pre- and post-treatment for drug efficacy/safety, the NiMS model indicates that compared to the pre-treatment status, the probability that the post-treatment status is maintained or improved equals the probability that the post-treatment status is deteriorated or maintained. On the other hand, the ENiMS and GNiMS models are useful to assess that whether the subsequent rating is maintaining or improving the previous one. The ENiMS model assumes that the treatment effect is always constant and independent of the pre-treatment status. Whereas, the GNiMS model assumes that the treatment effect depends on the pre-treatment status.

Additionally, we gave the decompositions of the NiMS and ENiMS models. These decompositions (i.e., Theorems 1 and 2) may help visualize the reason for a poor fit of the NiMS and ENiMS models.

Acknowledgements

The authors thank the editor and the anonymous referee for the valuable comments.

References

[1] Agresti, A. (1984) *Analysis of Ordinal Categorical Data*, Wiley, Hoboken, New Jersey.

[2] Agresti, A. (2019) *An Introduction to Categorical Data Analysis*, 3rd edition, Wiley, Hoboken, New Jersey.

[3] Akaike, H. (1974) A new look at the statistical model identification, *IEEE Transactions on Automatic Control* **19**, 716–723.

[4] Bowker, A. H. (1948) A test for symmetry in contingency tables, *Journal of the American Statistical Association* **43**, 572–574.

[5] Carlson, M., Vigen, C. L. P., Rubayi, S., Blanche, E. I., Blanchard, J., Atkins, M., Bates-Jensen, B., Garber, S. L., Pyatak, E. A., Diaz, J., Florindez, L. I., Hay, J. W., Mallinson, T., Unger, J. B., Azen, S. P., Scott, M., Cogan, A. and Clark, F. (2019) Lifestyle intervention for adults with spinal cord injury: Results of the USC-RLANRC Pressure Ulcer Prevention Study, *The Journal of Spinal Cord Medicine* **42**, 2–19.

[6] Caussinus, H. (1965) Contribution à l'analyse statistique des tableaux de corrélation, *Annales de la Faculté des Sciences de l'Université de Toulouse* **29**, 77–182.

[7] Clark, F. (2019) Pressure Ulcer Prevention Study. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor: ICPSR], 2019-10-04. https://doi.org/10.3886/E100595V2

[8] Goodman, L. A. (1979) Multiplicative models for square contingency tables with ordered categories, *Biometrika* **66**, 413–418

[9] Iki, K., Tahata, K. and Tomizawa, S. (2010) Decomposition of marginal homogeneity into logit and mean ridits equality for square contingency tables, *Journal of Mathematics and Statistics* **6**, 64–67.

[10] Konishi, S. and Kitagawa, G. (2008) *Information Criteria and Statistical Modeling*, Springer, New York.

[11] McCullagh, P. (1978) A class of parametric models for the analysis of square contingency tables with ordered categories, *Biometrika* **65**, 413–418.

[12] Read, C. B. (1977) Partitioning chi-square in contingency tables: A teaching approach, *Communications in Statistics-Theory and Methods* **6**, 553–562.

[13] Saigusa, Y., Maruyama, T., Tahata, K. and Tomizawa, S. (2018) Extended marginal homogeneity model based on complementary log-log transform for square tables, *International Journal of Statistics and Probability* **7**, 27–31.

[14] Schuette, W., Tesch, H., B ü ttner, H., Krause, T., Soldatenkova, V. and Stoffregen, C. (2012) Second-line Treatment of Stage III/IV Non-Small-Cell Lung Cancer (NSCLC) with pemetrexed in routine clinical practice: Evaluation of performance status and health-related quality of life, *BMC Cancer* **12**, 1–12.

[15] Stuart, A. (1955) A test for homogeneity of the marginal distributions in a two-way classification, *Biometrika* **42**, 412–416.

[16] Tahata, K. and Tomizawa, S. (2008) Generalized marginal homogeneity model and its relation to marginal equimoments for square contingency tables with ordered categories, *Advances in Data Analysis and Classification* **2**, 295–311.

[17] Tomizawa, S. (1984) Three kinds of decompositions for the conditional symmetry model in a square contingency table, *Journal of the Japan Statistical Society* **14**, 35–42.

[18] Tomizawa, S. (1991) Decomposing the marginal homogeneity model into two models for square contingency tables with ordered categories, *Calcutta Statistical Association Bulletin* **41**, 201–207.

[19] Tomizawa, S. (1995) A generalization of the marginal homogeneity model for square contingency tables with ordered categories, *Journal of Educational and Behavioral Statistics* **20**, 349–360.

[20] Yamamoto, K., Shinoda, S. and Tomizawa, S. (2011) Decompositions for ordinal quasi-symmetry model in square contingency tables with ordered categories, *Journal of Mathematics and Statistics* **7**, 314–318.

Appendix

We shall consider the MLE of the expected frequencies under each proposed model.

(i) NiMS model

To obtain the MLEs of the expected frequencies under the NiMS model, we must maximize the Lagrangian

$$L = \sum_{i=1}^{r}\sum_{j=1}^{r} n_{ij} \log p_{ij} - \lambda(\sum_{i=1}^{r}\sum_{j=1}^{r} p_{ij} - 1) - \sum_{i=1}^{r} \psi_i (W_{1(i)} - W_{2(i)})$$

with respect to $\{p_{ij}\}$, λ, and $\{\psi_i\}$. We obtain the partial derivatives of L as

$$\frac{\partial L}{\partial p_{st}} = \begin{cases} \dfrac{n_{st}}{p_{st}} - \lambda - \displaystyle\sum_{k=s}^{t} \psi_k & (s < t), \\ \dfrac{n_{st}}{p_{st}} - \lambda + \displaystyle\sum_{k=t}^{s} \psi_k & (s > t), \\ \dfrac{n_{st}}{p_{st}} - \lambda & (s = t), \end{cases}$$

for $s = 1, \ldots, r$ and $t = 1, \ldots, r$,

$$\frac{\partial L}{\partial \lambda} = -\sum_{i=1}^{r}\sum_{j=1}^{r} p_{ij} + 1,$$

and

$$\frac{\partial L}{\partial \psi_s} = -W_{1(s)} + W_{2(s)}$$

for $s = 1, \ldots, r$. The MLEs are obtained as solutions of equations for these partial derivatives of L that are equal to zero.

(ii) ENiMS model

Similarly, we must maximize the Lagrangian

$$L = \sum_{i=1}^{r}\sum_{j=1}^{r} n_{ij} \log p_{ij} - \lambda(\sum_{i=1}^{r}\sum_{j=1}^{r} p_{ij} - 1) - \sum_{i=1}^{r} \psi_i (W_{1(i)} - \Delta W_{2(i)})$$

with respect to $\{p_{ij}\}$, λ, $\{\psi_i\}$, and Δ. We obtain the partial derivatives of L

as

$$\frac{\partial L}{\partial p_{st}} = \begin{cases} \dfrac{n_{st}}{p_{st}} - \lambda - \displaystyle\sum_{k=s}^{t} \psi_k & (s < t), \\ \dfrac{n_{st}}{p_{st}} - \lambda + \Delta \displaystyle\sum_{k=t}^{s} \psi_k & (s > t), \\ \dfrac{n_{st}}{p_{st}} - \lambda - (1-\Delta)\psi_s & (s = t), \end{cases}$$

for $s = 1, \ldots, r$ and $t = 1, \ldots, r$,

$$\frac{\partial L}{\partial \lambda} = -\sum_{i=1}^{r}\sum_{j=1}^{r} p_{ij} + 1,$$

$$\frac{\partial L}{\partial \psi_s} = -W_{1(s)} + \Delta W_{2(s)}$$

for $s = 1, \ldots, r$ and

$$\frac{\partial L}{\partial \Delta} = \sum_{i=1}^{r} \psi_i W_{2(i)}.$$

The MLEs are obtained as solutions of equations for these partial derivatives of L that are equal to zero.

(iii) GNiMS model

Similarly, we must maximize the Lagrangian

$$L = \sum_{i=1}^{r}\sum_{j=1}^{r} n_{ij} \log p_{ij} - \lambda\Big(\sum_{i=1}^{r}\sum_{j=1}^{r} p_{ij} - 1\Big) - \sum_{i=1}^{r} \psi_i (W_{1(i)} - \Delta\Phi^{i-1} W_{2(i)})$$

with respect to $\{p_{ij}\}$, λ, $\{\psi_i\}$, Δ, and Φ. We obtain the partial derivatives of L as

$$\frac{\partial L}{\partial p_{st}} = \begin{cases} \dfrac{n_{st}}{p_{st}} - \lambda - \displaystyle\sum_{k=s}^{t} \psi_k & (s < t), \\ \dfrac{n_{st}}{p_{st}} - \lambda + \Delta \displaystyle\sum_{k=t}^{s} \Phi^{k-1}\psi_k & (s > t), \\ \dfrac{n_{st}}{p_{st}} - \lambda - (1 - \Delta\Phi^{s-1})\psi_s & (s = t), \end{cases}$$

for $s = 1, \ldots, r$ and $t = 1, \ldots, r$,

$$\frac{\partial L}{\partial \lambda} = -\sum_{i=1}^{r}\sum_{j=1}^{r} p_{ij} + 1,$$

$$\frac{\partial L}{\partial \psi_s} = -W_{1(s)} + \Delta \Phi^{s-1} W_{2(s)}$$

for $s = 1, \ldots, r$,

$$\frac{\partial L}{\partial \Delta} = \sum_{i=1}^{r} \Phi^{i-1} \psi_i W_{2(i)}$$

and

$$\frac{\partial L}{\partial \Phi} = \Delta \sum_{i=1}^{r} (i-1) \Phi^{i-2} \psi_i W_{2(i)}.$$

The MLEs are obtained as solutions of equations for these partial derivatives of L that are equal to zero.

Table 1: Shifts in the SF-36 general health subscale score 12 months from the baseline in patients with SCIs who received PUPP intervention. Data from Clark (2019).

SF-36 subscale score at baseline	SF-36 subscale score at 12 months				
	(1)	(2)	(3)	(4)	Total
(1)	2	3	2	1	8
	(2.00)	(2.49)	(1.76)	(0.77)	
(2)	1	8	6	5	20
	(1.26)	(8.00)	(5.24)	(3.81)	
(3)	2	4	17	5	28
	(2.32)	(4.68)	(17.00)	(4.55)	
(4)	1	3	3	10	17
	(1.44)	(4.37)	(3.33)	(10.00)	
Total	6	18	28	21	73

(1):0$-$ $\leq$ 25, (2):$>$ 25$-$ $\leq$ 50, (3):$>$ 50$-$ $\leq$ 75, (4):$>$ 75$-$ $\leq$ 100.
The parenthesized values are the MLEs of the expected frequencies under the NiMS model.

Table 2: Values of the likelihood ratio chi-square statistic G^2 and AIC^+ for each model applied to the data in Table 1.

Applied Model	For Table 1			
	df	G^2	p-value	AIC^+
NiMS	4	1.552	0.817	-6.448
MH	3	1.190	0.756	-4.810
ENiMS	3	0.612	0.894	-5.388
GNiMS	2	0.543	0.762	-3.457
ZM[g_1]	1	1.419	0.234	-0.581
CDE	1	0.074	0.786	-1.926

Table 3: Shift table presenting the number of patients by the baseline KI and KI after the second treatment cycle with pemetrexed. Data from Schuette et al. (2012).

KI at baseline	KI after second treatment cycle				
	(1)	(2)	(3)	(4)	Total
(1)	248	36	5	10	299
	(248.87)	(42.00)	(4.98)	(9.68)	
(2)	36	49	23	15	123
	(29.05)	(44.55)	(23.98)	(15.18)	
(3)	4	11	13	9	37
	(4.35)	(11.37)	(15.72)	(7.65)	
(4)	1	1	1	9	12
	(1.21)	(1.14)	(1.69)	(9.58)	
Total	289	97	42	43	471

(1):$\geq$ 80%, (2):70%, (3):60%, (4):$\leq$ 50%.
KI means Karnofsky Index.
The parenthesized values are the MLEs of the expected frequencies under the GNiMS model.

Table 4: Values of the likelihood ratio chi-square statistic G^2 and AIC$^+$ for each model applied to the data in Table 3.

Applied Model	For Table 3			
	df	G^2	p-value	AIC$^+$
NiMS	4	34.853*	< 0.001	26.853
MH	3	33.898*	< 0.001	27.898
ENiMS	3	34.785*	< 0.001	28.785
GNiMS	2	4.122	0.127	0.122
ZM[g_1]	1	17.905*	< 0.001	15.905
CDE	1	33.229*	< 0.001	31.229

* means significant at 0.05 level.

Mana Aizawa
Department of Information Sciences, Faculty of Science and Technology
Tokyo University of Science
Noda City, Chiba, 278-8510, Japan
E-mail: aizawa.m01@gmail.com

Shuji Ando
Department of Information and Computer Technology, Faculty of Engineering
Tokyo University of Science
Katsushika Ku, Tokyo, 125-8585, Japan
E-mail: shuji.ando@rs.tus.ac.jp

Kouji Tahata
Department of Information Sciences, Faculty of Science and Technology
Tokyo University of Science
Noda City, Chiba, 278-8510, Japan
E-mail: kouji_tahata@rs.tus.ac.jp

Sadao Tomizawa
Department of Information Sciences, Faculty of Science and Technology
Tokyo University of Science
Noda City, Chiba, 278-8510, Japan
E-mail: tomizawa@is.noda.tus.ac.jp