

MAIN PAPER

Matrix decomposition in meta-analysis for extraction of adverse event pattern and patient-level safety profile

Kentaro Matsuura¹  | Jun Tsuchida²  | Shuji Ando² | Takashi Sozu²

¹Department of Management Science, Graduate School of Engineering, Tokyo University of Science, Tokyo, Japan

²Department of Information and Computer Technology, Faculty of Engineering, Tokyo University of Science, Tokyo, Japan

Correspondence

Kentaro Matsuura, Department of Management Science, Graduate School of Engineering, Tokyo University of Science, 6-3-1 Nijjuku, Katsushika-ku, Tokyo, Japan.

Email: 4419703@ed.tus.ac.jp

Abstract

The purpose of assessing adverse events (AEs) in clinical studies is to evaluate what AE patterns are likely to occur during treatment. In contrast, it is difficult to specify which of these patterns occurs in each patient. To tackle this challenging issue, we constructed a new statistical model including nonnegative matrix factorization by incorporating background knowledge of AE-specific structures such as severity and drug mechanism of action. The model uses a meta-analysis framework for integrating data from multiple clinical studies because insufficient information is derived from a single trial. We demonstrated the proposed method by applying it to real data consisting of three Phase III studies, two mechanisms of action, five anticancer treatments, 3317 patients, 848 AE types, and 99,546 AEs. The extracted typical treatment-specific AE patterns coincided with medical knowledge. We also demonstrated patient-level safety profiles using the data of AEs that were observed by the end of the second cycle.

KEYWORDS

clinical trial, co-occurrence, nonnegative matrix factorization, normal dynamic linear model, pattern extraction

1 | INTRODUCTION

In order to use medications properly, it is imperative to assess the adverse events (AEs). This assessment is made concomitantly with efficacy evaluations in clinical studies. The purpose of assessing AEs in clinical studies is to evaluate the AE patterns that are likely to occur during treatment. The occurrences of AEs are summarized per treatment, and a meta-analysis is often applied to the summarized AE data.¹⁻⁴ However, such a population-level method does not predict the occurrences of AEs at the patient level, and thus the meta-analysis cannot be used to determine a patient-level treatment strategy. At the patient-level, a predictive model including Poisson- or logistic regression uses the occurrence of an AE as an outcome and data such as age, weight, and biomarkers as predictors.^{5,6} In general, however, hundreds of different AE types are observed in clinical trials, with low occurrence frequency. It is impractical to predict all AE types, even though the data of patients characteristics are available. At the patient level, an AE pattern that is likely to occur is sufficient information to support the treatment strategy, even though the occurrence of each AE type cannot be predicted. Therefore, it is important to know which AE pattern occurs at the patient level. We refer to the extent to which a patient has each pattern component as “patient-level safety profile”.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. Pharmaceutical Statistics published by John Wiley & Sons Ltd

To obtain AE patterns and patient-level profiles, this paper utilizes co-occurrence relationships of patient-level AEs. For example, among the AEs of anticancer treatments, “vomiting” and “nausea” tend to occur simultaneously. Statistical models dealing with such co-occurrences include matrix decomposition which could probably extract the “vomiting and nausea pattern.” In other fields, such as natural language processing, matrix decomposition with nonnegative constraints on the elements (nonnegative matrix factorization, NMF^{7,8}) is often applied to nonnegative data such as the number of occurrences. For example, one study applied NMF to the number of occurrences of words in web postings or product labels for drugs and was able to extract patterns.⁹ To the best of our knowledge, however, NMF has not yet been applied to patient-level AEs. Meanwhile, a previous study applied factor analysis, a form of matrix decomposition, to data from approximately 200 patients in a single clinical study and focused on fewer than 20 AEs of interest. Due to the small amount of data, however, only two patterns were extracted and were not used for construction of the patient-level safety profiles.¹⁰ Moreover, factor analysis contains strong constraints on the factor score matrix which can cause several issues. For example, the estimation results contain negative values, making the interpretation of patterns and safety profiles difficult.

The aim of this paper is to develop a method which identifies treatment-specific AE patterns and constructs patient-level safety profiles. We used patient AE information gathered from treatment onset to a predetermined time point. In this way, we could know the profiles and predict which AEs were likely to occur from the aforementioned time point to the end of treatment.

To this end, we intended to apply NMF to the AE data. However, this could not be realized without developing a new statistical model expanding NMF. Because AEs in clinical studies include both disease-specific and treatment-specific events, we needed to build a model to separate them. Moreover, AE data from clinical trials includes severity (ordinal categorical data) and the information of AEs varies with severity; hence, our model was constructed to avoid losing this information. Also, because a single clinical study does not contain sufficient data for our desired analysis, we included the assumption that treatment-specific AE patterns are similar for drugs with the same mechanism of action. This allowed us to use a sufficient number of patients from multiple clinical studies within a meta-analysis framework. These expansions enabled us to use actual data to extract AE patterns and demonstrate patient-level safety profiles.

In Section 2, an example of patient-level AEs data handled in this paper, our statistical model, and patient-level safety profile are described. In Section 3, the results from application of these methods are presented. In Section 4, we summarize and discuss our findings.

2 | DATA AND ANALYSIS METHODS

Although our method was widely applicable to clinical study data with patient-level AEs, specific data was explained earlier to understand the proposed model in Section 2.2.

2.1 | Data

Three datasets (registration numbers 118, 120, and 127) were chosen from the Project Data Sphere, which contains patient-level AE information. Each dataset corresponds to one Phase III study and has only control arm data. These datasets were selected because they were derived from three studies with the highest number of participants. They consisted of breast cancer patients who had undergone chemotherapy. The dosages and disease conditions slightly differed among the studies. Nevertheless, these discrepancies were not considered because we assumed that they had negligible influences on the analysis.

The datasets included AEs that occurred during baseline or treatment periods. Their severities were defined by the Common Terminology Criteria for Adverse Events (CTCAE, 1: mild, 2: moderate, 3: severe, 4: life-threatening, and 5: death) and ranged from 1 to 4. The baseline data represent AEs occurring during the pre-treatment period (from registration to randomization). This type of data is generally recorded. In this paper, because alopecia occurred in more than 70% of the patients, its pattern was clear and it was excluded from the analysis. Patients without AEs were also excluded, because their safety profiles were estimated to be zero and their data did not contribute to the estimation of AE patterns. There were 3317 patients and 99,546 AEs analyzed.

The outline of each arm included in the datasets is shown in Table 1. The index p is defined for the treatment periods to divide AEs by treatment. For example, a clinical study with $s = 1$ includes one arm named “AC \rightarrow T,” in which 1625 patients were enrolled and treated with four cycles of AC (doxorubicin + cyclophosphamide) followed by four cycles of T (docetaxel). $p = 1$ represents the baseline period, and the AEs occurring during this period are stored in the matrix Y_1 as described below. $p = 2$ represents the treatment period using AC, and the AEs that occurred during

TABLE 1 Summary of each arm and indices in the three datasets

| Study index s | Arm | Period index p | Treatment | Treatment index m | Mechanism index t | Number of patients | Number of cycles |
|-----------------|--------|------------------|-----------|---------------------|---------------------|--------------------|------------------|
| 1 | AC→T | 1 | Baseline | NA | NA | 1625 | NA |
| 1 | AC→T | 2 | AC | 1 | 1 | 1625 | 4 |
| 2 | AC→T | 3 | T | 5 | 2 | 1583 | 4 |
| 2 | FAC | 4 | FAC | 2 | 1 | 731 | 6 |
| 3 | A→CMF | 5 | Baseline | NA | NA | 480 | NA |
| 3 | A→CMF | 6 | A | 3 | 1 | 480 | 4 |
| 3 | A→CMF | 7 | CMF | 4 | 1 | 459 | 3 |
| 3 | AC→CMF | 8 | Baseline | NA | NA | 481 | NA |
| 3 | AC→CMF | 9 | AC | 1 | 1 | 481 | 4 |
| 3 | AC→CMF | 10 | CMF | 4 | 1 | 472 | 3 |

Note: s represents the index of the studies and p represents the index of the periods. A is doxorubicin, C is cyclophosphamide, F is 5-fluorouracil, M is methotrexate, and T is docetaxel. AC, FAC, and CMF represents combination therapies. t represents the index of the treatments and m represents the index of the mechanisms. Docetaxel is a taxane-based anticancer drug that binds to microtubules and inhibits mitosis. The other drugs inhibit DNA or RNA synthesis and are often used in combination. Their AEs are relatively similar. Therefore, two mechanisms of action were assumed here, namely, inhibition of nucleic acid synthesis and inhibition of mitosis. Cycles represent the lengths of the treatment periods.

this period are stored in \mathbf{Y}_2 . Similarly, $p = 3$ represents the treatment period using T. The number of patients decreases to 1583 in $p = 3$ because 42 patients dropped out of the study. We also define the treatment index and the mechanism index to facilitate the description of equations. The index of the periods was $p = 1, \dots, 10$, the index of the studies was $s \in \{1, 2, 3\}$, the index of the treatments was $t \in \{1, \dots, 5\}$ and the index of the mechanisms of action was $m \in \{1, 2\}$. The study and the treatment were determined by period p and the mechanism of action was determined by treatment t . Therefore, they were suffixed such as s_p , t_p , and m_{t_p} when they were accessed using p . For example, $s_3 = 1$, $t_3 = 5$ and $m_5 = 2$.

In this paper, AEs with the same name but different severities were distinguished. The combination of name and severity was treated as “AE type”. For example, VOMIT with severity 2 is “VOMIT_2.” Consequently, the total number of AE types was $J = 848$.

These datasets included the number of occurrences of each AE type per cycle. The number of occurrences was summarized per period p and the result was expressed as $N_p \times J$ matrix \mathbf{Y}_p ($p = 1, \dots, 10$), where N_p is the number of patients who participated in period p . \mathbf{Y}_p was a sparse matrix in which many elements were 0. $Y_{ij}^{(p)}$, which is the (i, j) element of \mathbf{Y}_p , represents the number of occurrences of AE type j for patient i during period p .

2.2 | Analysis methods

2.2.1 | Statistical model

Here, the statistical model for generating \mathbf{Y}_p ($p = 1, \dots, 10$) is explained. Scalars are in italics, vectors are in bold font, and matrices are in bold italics.

Hereafter, one period p was selected and fixed. \mathbf{Y}_p was written as \mathbf{Y} , N_p as N , s_p as s , t_p as t , and m_{t_p} as m . The patient index was written as i and the AE type index was written as j .

Because \mathbf{Y} was count data, it was assumed that each element of \mathbf{Y} independently followed the following Poisson distribution:

$$Y_{ij} \sim \text{Poisson}(C_i \lambda_{ij}) \quad i = 1, \dots, N \quad j = 1, \dots, J \quad (1)$$

where C_i is a constant representing the length of the treatment period of patient i (i.e., the number of cycles in the aforementioned data).

λ_{ij} was separated into $\beta_{ij}^{(t)}$ derived from treatment t and $\alpha_{ij}^{(s)}$ derived from study s . In other words,

$$\lambda_{ij} = \beta_{ij}^{(t)} + \alpha_{ij}^{(s)} \quad i = 1, \dots, N \quad j = 1, \dots, J \quad (2)$$

where $\alpha_{ij}^{(s)}$ is a parameter per study representing the occurrence intensity of each AE caused by disease, radiotherapy, etc. For the baseline data, only $\alpha_{ij}^{(s)}$ was considered and it was assumed that $\beta_{ij}^{(t)} = 0$.

$\beta_{ij}^{(t)}$ is the (i, j) element of the $N \times J$ matrix $\beta^{(t)}$, and $\alpha_{ij}^{(s)}$ is the (i, j) element of the $N \times J$ matrix $\alpha^{(s)}$. The occurrence patterns of AEs are usually limited.¹¹ Therefore, it was assumed that they consisted of a combination of relatively few patterns with little information loss and the matrix decomposition was applied as follows:

$$\beta^{(t)} = \theta^{(t)} \phi^{(t)} \quad (3)$$

$$\alpha^{(s)} = \xi^{(s)} \eta^{(s)} \quad (4)$$

where $\theta^{(t)}$ is an $N \times K$ matrix, $\phi^{(t)}$ is a $K \times J$ matrix, $\xi^{(s)}$ is an $N \times L$ matrix, and $\eta^{(s)}$ is an $L \times J$ matrix. K and L are $\ll J$ and represent the numbers of AE type occurrence patterns. ϕ and η are common to all patients and represent the occurrence intensities of all AE types determined for each pattern. θ and ξ represent the extent to which each patient has each pattern component. Expressing a matrix by the product of low-rank matrices generally facilitates interpretation and improves prediction accuracy by eliminating noise.

The decomposition of β into θ and ϕ is not unique. The constraints for θ and ϕ make them identifiable except the order of the components and make the estimation possible. For example, the constraint that θ and ϕ are orthogonal matrices makes matrix decomposition equivalent to principal component analysis. However, the approximation of \mathbf{Y} is poor because orthogonality is a strong constraint. It is difficult to interpret the result because certain elements of θ and ϕ are negative. Moreover, some λ_{ij} , which represents the average number of occurrences, may also be negative. Therefore, in this paper, the constraint that all elements of θ and ϕ are non-negative was added (NMF).

$\theta^{(t)}$ was reparametrized as follows:

$$\theta^{(t)} = \exp \tilde{\theta}^{(t)} \quad (5)$$

where $\tilde{\theta}^{(t)}$ is an $N \times K$ matrix, and the exp function operates on each element and returns a matrix.

For $\phi^{(t)}$, background knowledge of treatments was used, and it was assumed that $\phi^{(t)}$ is constructed as follows:

$$\Phi_k^{(t)} = \text{softmax}(\mathbf{M}_k^{(m)} + \mathbf{R}_k^{(t)}) \quad k = 1, \dots, K \quad (6)$$

where $\Phi_k^{(t)}$ represents the row vector of the k -th row of $\phi^{(t)}$. $\mathbf{u} = \text{softmax}(\mathbf{v})$ converts vector \mathbf{v} into vector \mathbf{u} whose elements are nonnegative and have a sum of 1, using the following formula:

$$u_j = \frac{\exp v_j}{\sum_j \exp v_j}$$

Equation (6) facilitates comparisons between treatments because the sum of J elements of each row (pattern) of $\phi^{(t)}$ becomes 1. $\mathbf{M}^{(m)}$ is a $K \times J$ matrix representing the mean effect of the mechanism of action m . $\mathbf{R}^{(t)}$ is a $K \times J$ matrix representing the random effect of treatment t . $\mathbf{M}^{(1)}$ and $\mathbf{M}^{(2)}$ were considered for all data. For “AC,” “FAC,” “A,” and “CMF,” which inhibit nucleic acid synthesis, the random effects of $\mathbf{R}^{(1)}$, $\mathbf{R}^{(2)}$, $\mathbf{R}^{(3)}$, and $\mathbf{R}^{(4)}$ were considered, respectively (see Table 1).

Similar constraints were assumed for $\alpha^{(s)}$ as in $\beta^{(t)}$. $\xi^{(s)}$ was reparametrized as follows to make each element nonnegative:

$$\xi^{(s)} = \exp \tilde{\xi}^{(s)} \quad (7)$$

$\eta^{(s)}$ was defined using $L \times J$ matrix $\mathbf{S}^{(s)}$ as follows:

$$\boldsymbol{\eta}_l^{(s)} = \text{softmax}(\mathbf{S}_l^{(s)}) \quad l = 1, \dots, L \quad (8)$$

If AEs caused by disease were of interest, $\mathbf{S}_l^{(s)}$ could be divided into mean- and random effects as for the treatments.

In an occurrence pattern, AE types with the same name and similar severities (e.g., NAUSEA_2 and NAUSEA_3) should have similar occurrence intensities. Therefore, $\mathbf{M}^{(m)}$, $\mathbf{R}^{(t)}$ and $\mathbf{S}^{(s)}$ were constrained using a NDLM (normal dynamic linear model),¹² which is often applied in dose-response curves:

$$M_{kj'}^{(m)} \sim \mathcal{N}(M_{kj}^{(m)}, \sigma_G^2) \quad k = 1, \dots, K \quad (j, j') \in \mathcal{J} \quad (9)$$

$$R_{kj'}^{(t)} \sim \mathcal{N}(R_{kj}^{(t)}, \sigma_G^2) \quad k = 1, \dots, K \quad (j, j') \in \mathcal{J} \quad (10)$$

$$S_{lj'}^{(s)} \sim \mathcal{N}(S_{lj}^{(s)}, \sigma_G^2) \quad l = 1, \dots, L \quad (j, j') \in \mathcal{J} \quad (11)$$

where \mathcal{J} is a set of the combination of AE types j and j' satisfying the conditions that j and j' have the same AE name, their severities are adjacent, and $j < j'$.

The addition of a constant to each element of the argument vector does not change the output of the softmax function. Therefore, $\mathbf{M}^{(m)}$, $\mathbf{R}^{(t)}$ and $\mathbf{S}^{(s)}$ were constrained as follows so that they could be identified and estimated:

$$M_{kj}^{(m)} \sim \mathcal{N}(0, \sigma_M^2) \quad k = 1, \dots, K \quad j \in \mathfrak{J} \quad (12)$$

$$R_{kj}^{(t)} \sim \mathcal{N}(0, \sigma_R^2) \quad k = 1, \dots, K \quad j \in \mathfrak{J} \quad (13)$$

$$S_{lj}^{(s)} \sim \mathcal{N}(0, \sigma_S^2) \quad l = 1, \dots, L \quad j \in \mathfrak{J} \quad (14)$$

where \mathfrak{J} is a set of AE type j satisfying the condition that j has the lowest severity in each AE name.

Considering Equations (1)-(14) for all data \mathbf{Y}_p ($p = 1, \dots, 10$), the log of the posterior probability was constructed as follows:

$$\begin{aligned} & \log P(\tilde{\boldsymbol{\theta}}^{(t)}, \mathbf{M}^{(m)}, \mathbf{R}^{(t)}, \tilde{\boldsymbol{\xi}}^{(s)}, \mathbf{S}^{(s)} \mid \mathbf{Y}_p, \sigma_M, \sigma_R, \sigma_S, \sigma_G) \\ &= \sum_{p=1}^{10} \sum_{i=1}^{N_p} \sum_{j=1}^J \log \text{Poisson} \left(Y_{ij}^{(p)} \mid C_i^{(p)} \left[\exp \tilde{\theta}^{(t_p)} \text{softmax}(\mathbf{M}^{(m_{t_p})} + \mathbf{R}^{(t_p)}) + \exp \tilde{\xi}^{(s_p)} \text{softmax}(\mathbf{S}^{(s_p)}) \right]_{ij} \right) \\ &+ \sum_{m=1}^2 \sum_{k=1}^K \left[\sum_{j \in \mathfrak{J}} \log \mathcal{N}(M_{kj}^{(m)} \mid 0, \sigma_M^2) + \sum_{(j, j') \in \mathcal{J}} \log \mathcal{N}(M_{kj'}^{(m)} \mid M_{kj}^{(m)}, \sigma_G^2) \right] \\ &+ \sum_{t=1}^5 \sum_{k=1}^K \left[\sum_{j \in \mathfrak{J}} \log \mathcal{N}(R_{kj}^{(t)} \mid 0, \sigma_R^2) + \sum_{(j, j') \in \mathcal{J}} \log \mathcal{N}(R_{kj'}^{(t)} \mid R_{kj}^{(t)}, \sigma_G^2) \right] \\ &+ \sum_{s=1}^3 \sum_{l=1}^L \left[\sum_{j \in \mathfrak{J}} \log \mathcal{N}(S_{lj}^{(s)} \mid 0, \sigma_S^2) + \sum_{(j, j') \in \mathcal{J}} \log \mathcal{N}(S_{lj'}^{(s)} \mid S_{lj}^{(s)}, \sigma_G^2) \right] \end{aligned} \quad (15)$$

where the softmax function operates on each row of a matrix and returns a matrix, $Poisson(y | \lambda)$ represents the probability mass at y of the Poisson distribution with parameter λ , and $\mathcal{N}(y | \mu, \sigma^2)$ represents the probability density at y of the normal distribution with parameters μ and σ^2 . In the Appendix A, we explain that the models and Equation (15) can be simplified under certain conditions.

2.2.2 | Estimation

$\theta^{(t)}$, $\mathbf{M}^{(m)}$, $\mathbf{R}^{(t)}$, $\xi^{(s)}$, and $\mathbf{S}^{(s)}$ were estimated by maximizing the log-probability $\log P$ in Equation (15). σ_G , σ_M , σ_R , and σ_S could also be estimated if there was sufficient data. In this paper, they were regarded as hyperparameters and assigned fixed values. Although the number of parameters is large, the parameters are identifiable because matrix decomposition is an informative constraint,⁷ and Equations (6), (8)–(14) do not cause the loss of identifiability.

The software used for this estimation were Stan 2.17.¹³ Maximum a posteriori (MAP) estimation was performed using the limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm. The Stan and R¹⁴ programs are available on GitHub,¹⁵ and statisticians can use our method by modifying parts of the programs.

2.2.3 | Hyperparameter determination

The number of patterns K and L were hyperparameters that had to be fixed before estimation. Although there are several determination methods, the technique applied here ensured hyperparameter interpretation according to background information. If the number of patterns was too small, several were combined and extracted as one pattern. As a rule, however, this manipulation should be avoided. Conversely, if the number of patterns was large, similar patterns were extracted.

Other determinations were based either on predictive performance or information criteria, such as the Akaike Information Criterion (AIC) and the Widely Applicable Information Criterion (WAIC).¹⁶ On the other hand, one study indicated that results derived from the number of patterns determined by these methods do not align with the actual interpretation.¹⁷

In this paper, the hyperparameters were determined as follows. The small value $L = 2$ was selected because there are far fewer AEs associated with the disease itself than those originating from anticancer treatments used in the clinical studies. The number of detectable patterns associated with the disease was assumed to be small. In fact, in the data in Section 2.1, less than 10% of the total AEs could be attributed to the disease. $K = 15$ was set because similar patterns occur when K is assigned higher values. For actual clinical studies, L and K must be determined after trying several values and discussing the results.

$\sigma_G = 1.5$ because the number of severities was as low as 1 to 4. $\sigma_M = 5$ and $\sigma_S = 5$ were sufficiently large to express any pattern and $\sigma_R = 0.5$ because it was assumed that the magnitude of the random effect was about 10% that of the mean effect.

2.3 | AE patterns

The estimate of $\phi^{(t)}$ is a set of AE type patterns specific to treatment t . $\Phi_k^{(t)}$ is pattern k and a vector of occurrence intensities of all AE types. AE types most likely to occur in pattern k are readily confirmed by rearranging its intensities in descending order and plotting them in a table or graph. Patterns where severe AEs are most likely to occur can also be confirmed.

The estimate of $\theta_i^{(t)}$ represents responsiveness of patient i to treatment t , and it provides condensed information of \mathbf{Y}_i through the patterns. $\theta_{ik}^{(t)}$ is the extent to which patient i has a pattern k component. $\theta^{(t)}$ can also be interpreted as the soft clustering result for the patients, where each patient can have components belonging to multiple clusters. The number of patients with high values of a pattern k component can be established by visualizing the distribution of $\theta_{\cdot,k}^{(t)}$, which represents the k -th column of $\theta^{(t)}$ (a column vector).

2.4 | Patient-level safety profile

Suppose that patient i' receives treatment t and his/her AEs (i.e., $\mathbf{Y}_{i'}$) are assessed to estimate $\boldsymbol{\theta}_{i'}^{(t)}$ at a particular time point in a new clinical study.

All parameters can be estimated from the data, including the AEs of patient i' obtained from the start of the treatment to the specified time point and the AEs from other patients and other clinical studies. The estimation will reveal which pattern components are abundant in patient i' and predict which AE types are most likely to occur from the specified time point to the end of treatment. This method is effective if computation time and resources suffice. However, it is not realistic to repeat parameter estimation whenever the assessment time point changes.

When prompt predictions are needed during a new clinical study, diverting the already estimated $\boldsymbol{\phi}^{(t)}$ (e.g., our estimated $\boldsymbol{\phi}^{(t)}$) enables rapid $\boldsymbol{\theta}_{i'}^{(t)}$ estimation. Assuming that $\boldsymbol{\phi}^{(t)}$ are sufficiently close to the true values and the influence of each study (i.e., $\alpha_{i'j}^{(s)}$) is sufficiently small, we maximize the following log-probability to estimate $\boldsymbol{\theta}_{i'}^{(t)}$:

$$\sum_{j=1}^J \log \text{Poisson} \left(Y_{i'j} \mid C \left[\sum_{k=1}^K \left(\exp \tilde{\theta}_{i'k}^{(t)} \right) \phi_{kj}^{(t)} \right] \right) \quad (16)$$

where C is a constant representing the length of the treatment period of patient i' . The posterior distribution may also be estimated by setting noninformative prior distributions or appropriate prior distributions for $\boldsymbol{\theta}_{i'}^{(t)}$.

The estimate of $\boldsymbol{\theta}_{i'}^{(t)}$ and Equations (2)–(4) indicate that $\lambda_{i'j}$ can be calculated. $\lambda_{i'j}$ is the mean number of occurrences of AE type j in a future treatment period. AE types with high $\lambda_{i'j}$ must be carefully monitored.

3 | RESULTS

We applied the method described in Section 2.2 to the data in Section 2.1 for AE pattern extraction and patient-level safety profile estimation. The results corresponding to Sections 2.3 and 2.4 are explained below.

3.1 | AE patterns

Here, the $\boldsymbol{\phi}^{(t)}$ patterns of treatment AC ($t = 1$) are shown. For the patterns of all treatments, see Table S1.

The 15 patterns of $\boldsymbol{\phi}^{(1)}$ were numbered in descending order of the log-probability increase. The results of patterns 1 to 15 are shown in Table 2. Because the log-probability increase involving patterns 1 to 8 was greater than 80% of that for all patterns, we focused on the results from patterns 1 to 8.

For instance, pattern 2 is a “nausea and vomiting pattern” in which nausea and vomiting of severities 2 and 3, respectively, tended to occur simultaneously. However, nausea and vomiting of severity 1 were segregated into patterns 1 and 5 because the numbers of their occurrences were very large. Pattern 4 was interpreted as a “constipation pattern” and included constipation, stomatitis, and weight increase. Pattern 6 was the “neuropathy pattern” involving taste perversion, diarrhea, insomnia, and lacrimation disorder. Pattern 7 was another “neuropathy pattern” with nausea of severity 2, nail disorder, vasodilation, anorexia, and weight decrease.

Next, we confirmed the responsiveness of the 2106 patients being administered AC. The distributions of $\boldsymbol{\theta}_{:k}^{(1)}$ corresponding to patterns 1 to 8 are shown in Figure 1. The $\boldsymbol{\theta}_{:k}^{(1)}$ for most patients was almost 0 but $\boldsymbol{\theta}_{:k}^{(1)}$ was high in a few cases.

For the remaining treatments, we can also extract 15 patterns $\boldsymbol{\phi}^{(t)}$ and plot distributions of $\boldsymbol{\theta}_{:k}^{(1)}$.

3.2 | Patient-level safety profile

We demonstrated patient-level safety profiles using the method in Section 2.4. Suppose that four new patients ($i' = 1', \dots, 4'$) received AC in a new clinical study. Fixing $\boldsymbol{\phi}^{(1)}$ to the values in Section 3.1 and applying Equation (16), we estimated $\boldsymbol{\theta}_{i'}^{(1)}$ by MAP estimation every time an AE occurred. $\boldsymbol{\theta}_{i'}^{(1)}$ can be calculated for any AE data. For example, the AE

TABLE 2 Result of $\phi^{(1)}$

| Pattern k | AE type | $\phi_{k,j}^{(1)}$ |
|-------------|---------------------|--------------------|
| 1 | NAUSEA_1 | 99.9 |
| 2 | VOMIT_2 | 44.7 |
| | NAUSEA_2 | 42.4 |
| | VOMIT_3 | 6.2 |
| | NAUSEA_3 | 6.0 |
| 3 | ASTHENIA_1 | 99.9 |
| 4 | CONSTIP_1 | 53.7 |
| | STOMATITIS_2 | 23.6 |
| | WEIGHT INC_1 | 17.6 |
| | WEIGHT INC_2 | 1.8 |
| | INFECT_2 | 1.3 |
| | STOMATITIS_3 | 0.6 |
| 5 | VOMIT_1 | 99.8 |
| 6 | TASTE PERVERS_1 | 22.5 |
| | DIARRHEA_1 | 19.4 |
| | INSOMNIA_1 | 16.8 |
| | LACRIMATION DIS_1 | 13.9 |
| | MENS DIS_1 | 13.2 |
| | SKIN DRY_1 | 4.7 |
| | FEVER_1 | 3.3 |
| | PHARYNGITIS_1 | 1.9 |
| | STOMATITIS_3 | 0.1 |
| 7 | NAUSEA_2 | 46.1 |
| | NAIL DIS_1 | 24.6 |
| | VASODILAT_2 | 13.9 |
| | ANOREXIA_2 | 6.3 |
| | PAIN ABDO_2 | 2.7 |
| | NAIL DIS_2 | 1.7 |
| | WEIGHT DEC_1 | 1.4 |
| | ANOREXIA_3 | 0.3 |
| | WEIGHT DEC_3 | 0.1 |
| 8 | STOMATITIS_1 | 99.3 |
| | STOMATITIS_3 | 0.3 |
| 9 | ASTHENIA_2 | 95.6 |
| | ASTHENIA_3 | 4.1 |
| 10 | HEADACHE_1 | 39.8 |
| | MYALGIA_1 | 20.7 |
| | RHINITIS_1 | 19.1 |
| | ARTHRALGIA_1 | 13.5 |
| | SKIN DIS_1 | 3.5 |
| | DIARRHEA_3 | 0.2 |
| Pattern k | AE type | $\phi_{k,j}^{(1)}$ |
| 11 | VASODILAT_1 | 30.6 |
| | ANOREXIA_1 | 29.9 |
| | PAIN_1 | 19.5 |
| | CONSTIP_2 | 13.4 |

(Continues)

TABLE 2 (Continued)

| Pattern k | AE type | $\phi_{kj}^{(1)}$ |
|-------------|-----------------------|-------------------|
| | DEPRESSION_1 | 2.6 |
| | CONJUNCTIVITIS_1 | 1.2 |
| | CONSTIP_3 | 0.2 |
| 12 | NEUROPATHY_1 | 17.6 |
| | DYSPNEA_2 | 10.3 |
| | RASH MAC PAP_1 | 9.9 |
| | EDEMA PERIPH_1 | 8.1 |
| | INSOMNIA_2 | 7.0 |
| | DIZZINESS_1 | 6.7 |
| | INFECT_1 | 4.7 |
| | TASTE PERVERS_2 | 2.9 |
| | DYSPNEA_3 | 0.6 |
| | HEADACHE_3 | 0.2 |
| 13 | MENS DIS_3 | 10.0 |
| | PAIN ABDO_1 | 6.9 |
| | INFECT_3 | 5.0 |
| | DIARRHEA_2 | 4.3 |
| | PAIN BACK_1 | 3.8 |
| | WEIGHT DEC_1 | 3.4 |
| | SWEAT_1 | 3.4 |
| | ARTHRALGIA_2 | 2.9 |
| 14 | DYSPEPSIA_1 | 49.1 |
| | HEADACHE_2 | 21.2 |
| | DYSPEPSIA_2 | 15.9 |
| | INJECT SITE REACT_1 | 8.7 |
| | HEADACHE_1 | 1.6 |
| | HEADACHE_3 | 0.5 |
| | DYSPEPSIA_3 | 0.4 |
| 15 | CONJUNCTIVITIS_1 | 10.9 |
| | COUGH INC_1 | 10.1 |
| | ANXIETY_1 | 9.5 |
| | DRY MOUTH_1 | 7.5 |
| | PAIN BONE_1 | 4.7 |
| | DRY EYE_1 | 4.4 |
| | ALLERG REACT_1 | 3.9 |
| | PALPITAT_1 | 3.7 |
| | BRONCHITIS_3 | 0.3 |
| | ALLERG REACT_3 | 0.3 |

Note: The unit of occurrence intensity was converted to % because the sum of occurrence intensities per pattern was set to 1 by the softmax function. The top eight AE types whose occurrence intensities are more than 1% are presented for each pattern. We also show the top two AE types with severity 3 or higher and whose occurrence intensities are more than 0.1% (bold letters).

data of the four patients were assumed to be the same as $i = 548,338,871,530$, who received AC, respectively (see Data S1).

The estimated $\theta_i^{(1)}$ at the end of the second cycle is shown in Figure 2 (left). In patient 1', pattern 7 component (i.e., $\theta_{17}^{(1)}$) was very high due to vasodilation of severity 2 and abdominal pain of severity 2. Abdominal pain of severity 2 was a rare AE type that occurred in only 1.9% (40 out of 2106) of patients who received treatment AC, contributing to

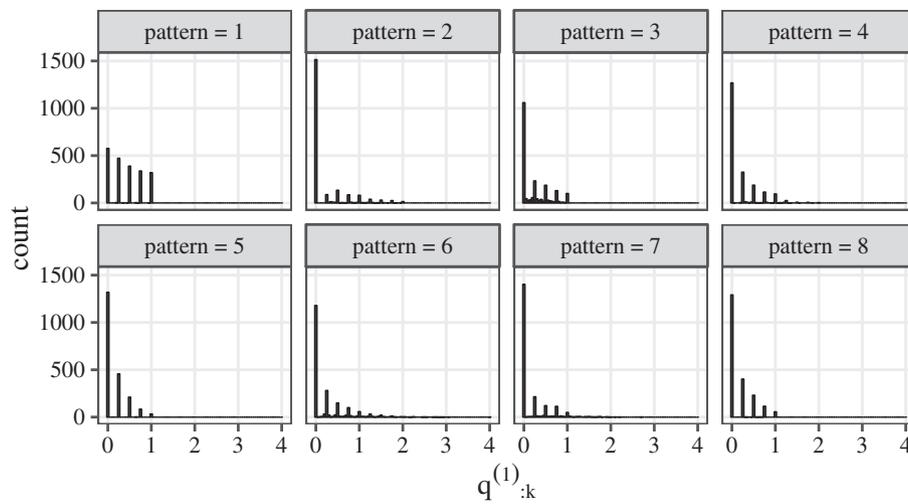


FIGURE 1 Distributions of $\theta_{i,k}^{(1)}$.
 Note: The distribution was partially discrete because the number of occurrences of AEs and the treatment periods (that is, the number of cycles) were integers here

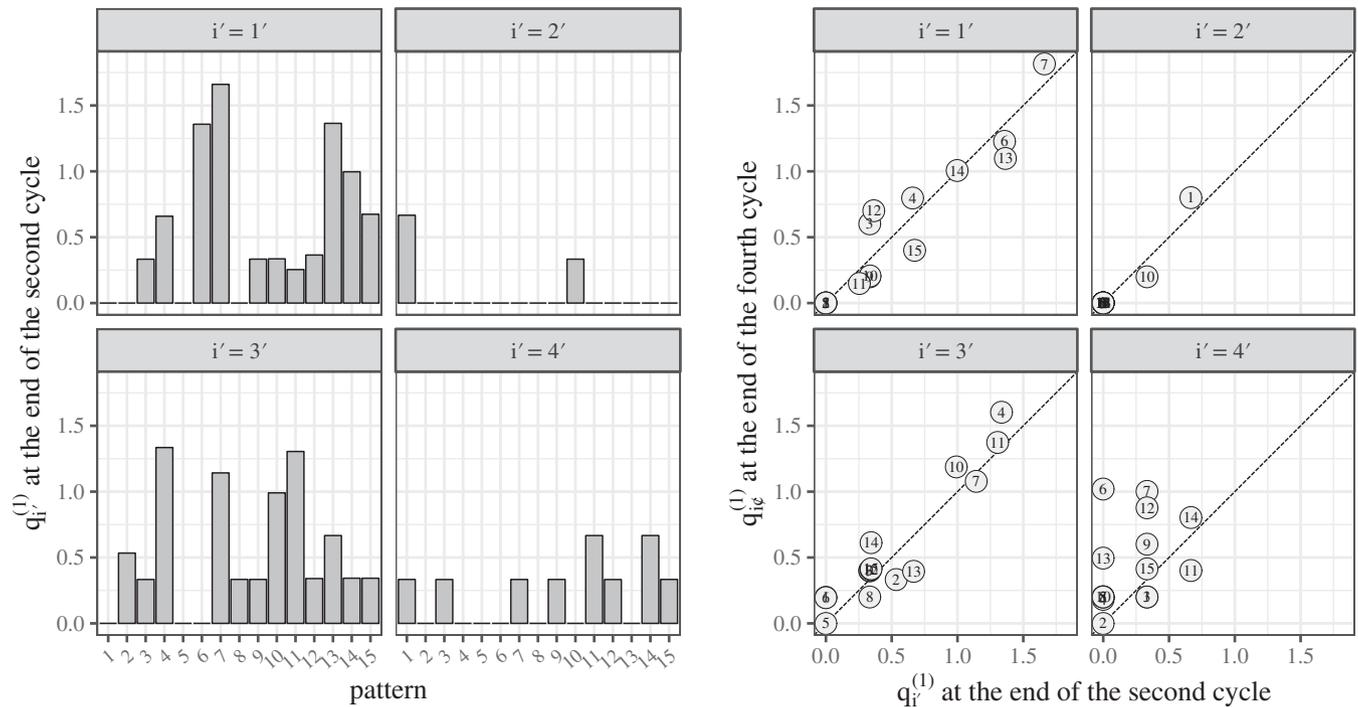


FIGURE 2 Estimated $\theta_{i'}^{(1)}$ of the four patients

an increase in $\theta_{1'7}^{(1)}$ because it is a pattern-7-specific AE type. Patient 2' had only two pattern components. The pattern components of patient 3' were separated into high and low. A scatter plot of the estimated $\theta_{i'}^{(1)}$ at the end of the second cycle and at the end of the fourth cycle is shown in Figure 2 (right). In patients 1' to 3', the points are near the diagonal, which suggests that $\theta_{i'}^{(1)}$ of patients 1' to 3' are stable. In contrast, because the points of patient $i' = 4'$ are far from the diagonal, $\theta_{4'}^{(1)}$ may remain unstable due to the insufficient AE data.

We assumed that the estimated $\theta_{i'}^{(1)}$ was close enough to the true value at the end of the second cycle. We then calculated $\lambda_{1'j}^{(1)}$ according to Equations (2) and (3) and listed the AE types most likely to occur in the future (Table 3). Those AE types should be carefully monitored.

| AE type | $\lambda_{1j}^{(1)}$ |
|-----------------|----------------------|
| NAUSEA_2 | 1.02 |
| DYSPEPSIA_1 | 0.61 |
| NAIL DIS_1 | 0.55 |
| CONSTIP_1 | 0.53 |
| TASTE PERVERS_1 | 0.33 |
| ASTHENIA_1 | 0.32 |
| VASODILAT_2 | 0.31 |
| DIARRHEA_1 | 0.29 |
| HEADACHE_2 | 0.26 |
| INSOMNIA_1 | 0.25 |
| MENS DIS_3 | 0.091 |
| INFECT_3 | 0.046 |
| FEVER_3 | 0.013 |
| ASTHENIA_3 | 0.012 |
| HEADACHE_3 | 0.008 |
| ANOREXIA_3 | 0.008 |
| LEUKOPENIA_3 | 0.008 |
| STOMATITIS_3 | 0.007 |
| LEUKOPENIA_4 | 0.006 |
| DIARRHEA_3 | 0.006 |

TABLE 3 Top 10 AE types and top 10 severe AE types with the largest $\lambda_{1j}^{(1)}$

4 | DISCUSSION

Our statistical model with patient-level AE data extracted AE patterns specific for each treatment. We clarified that our method allowed interpretable patient-level safety profiling in a clinical study such as the illustration of the data analysis in this paper. Our statistical model can use the number of occurrences of each AE in a patient to predict future AEs if the total number of occurrences is sufficiently large so that the estimated θ is close enough to the true value. The proposed method is also widely applicable to data obtained from clinical studies with patient-level AEs, regardless of disease or treatment. We believe that this paper is of great significance to the current situation where patient-level data is being released.

In Section 2.1, we excluded alopecia due to its common occurrence in patients. Such a frequent AE is called “stopwords” in the field of natural language processing (e.g., prepositions). It is common to remove such terms before matrix decomposition.¹⁸ If alopecia were included, many patterns would have contained alopecia, which would prevent pattern extraction and pattern interpretation. Therefore, we decided to exclude alopecia from the analysis in this paper.

A limitation that was noted in our method is that K must be determined before analysis (Section 2.2.3). For example, if we used $K = 10$, instead of pattern 6 and pattern 7, a mixture of those patterns would have been extracted. We used $K = 15$ after discussion because we thought they can be separated. Although non-parametric Bayesian methods can estimate K , further research is needed to examine whether it is effective for our model.

In our model, $\theta_{ik}^{(t)}$ represents the responsiveness of patient i to treatment t . If $\theta_{ik}^{(t)}$ and $\theta_{jk}^{(t)}$ are strongly correlated, a patient may be advised to opt for alternate treatments wherein severe AEs seldom occur based on the correlation. In this paper, however, there was no strong correlation between treatments. For this reason, it was difficult to predict AEs caused by a certain treatment based on the AEs induced by a different one.

For the sake of simplicity, we assumed that (1) the dosages of each drug were the same across clinical studies, (2) treatments were not changed within a cycle, and (3) the order in which the AEs occurred was insignificant.

Regarding (1), in practice, however, the dosages varied depending on the clinical studies. Regarding (2), treatments may change within a cycle due to drug withdrawal or dose reduction. (1) and (2) can be resolved by incorporating a dose–response model whose shape parameters are adjusted by the dosages actually administered. Regarding (3), the number, order, and timing of occurrences may all be important. This obstacle may be overcome by modeling the occurrence data as time series without summarizing them for each treatment period.

In this paper, we did not show the result of using patient information such as age or weight, however, these factors could be incorporated into our model. Because $\theta_{ik}^{(t)}$ represents the responsiveness of patient i to treatment t , $\theta_{ik}^{(t)}$ could be expressed using patient information as follows:

$$\begin{aligned}\mu_{ik} &= f(\mathbf{x}_i) \\ \theta_{ik}^{(t)} &\sim \text{LogNormal}(\mu_{ik}, \sigma^2)\end{aligned}$$

where \mathbf{x}_i is a vector of patient information for patient i . μ_{ik} is determined by a certain function f of \mathbf{x} , and $\theta_{ik}^{(t)}$ follows a lognormal distribution with μ_{ik} as a parameter. We can also incorporate exposure and other pharmacokinetics data into the model. In our study, we attempted to use age, weight, and laboratory measurements as \mathbf{x}_i for the data. The result showed that these information were not useful predictors for $\theta_{ik}^{(t)}$ in this data. In future research, the effective integration of patient information into the model will be investigated.

Performance evaluation using artificial data is impractical due to the complexity of the AE data structure. Our method should be validated by applying it to larger datasets to use in a clinical study.

ACKNOWLEDGMENTS

The sample study data presented in this paper were derived from www.projectdatasphere.org, which is maintained by Project Data Sphere, LLC. Neither Project Data Sphere, LLC nor the owners of any data have contributed to, approved or are in any way responsible for the contents of this analysis. Finally, we thank the editor and anonymous reviewers for their constructive comments, which helped us to improve the manuscript.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The original data cannot be shared due to the rules of Project Data Sphere. The data on results and analysis methods are available in the supplementary material section and GitHub.¹⁵

ORCID

Kentaro Matsuura  <https://orcid.org/0000-0001-5262-055X>

Jun Tsuchida  <https://orcid.org/0000-0003-2349-2709>

REFERENCES

1. Bradburn MJ, Deeks JJ, Berlin JA, Russell LA. Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events. *Stat Med*. 2007;26(1):53-77.
2. Crowe BJ, XHA, Berlin JA, et al. Recommendations for safety planning, data collection, evaluation and reporting during drug, biologic and vaccine development: a report of the safety planning, evaluation, and reporting team. *Clin Trials*. 2009;6(5):430-440.
3. Berry SM, Berry DA. Accounting for multiplicities in assessing drug safety: a three-level hierarchical mixture model. *Biometrics*. 2004;60(2):418-426.
4. Odani M, Fukimbara S, Sato T. A Bayesian meta-analytic approach for safety signal detection in randomized clinical trials. *Clin Trials*. 2017;14(2):192-200.
5. Green SM, Roback MG, Krauss B, et al. Predictors of airway and respiratory adverse events with ketamine sedation in the emergency department: an individual-patient data meta-analysis of 8,282 children. *Ann Emerg Med*. 2009;54(2):158-168.
6. von Dadelszen P, Payne B, Li J, et al. Prediction of adverse maternal outcomes in pre-eclampsia: development and validation of the FulPIERS model. *Lancet*. 2011;377(9761):219-227.
7. Paatero P, Tapper U. Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*. 1994;5(2):111-126.
8. Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. *Nature*. 1999;401(6755):788.

9. Wang Y, Gunashekar DR, Adam TJ, Zhang R. Mining adverse events of dietary supplements from product labels by topic modeling. *Stud Health Technol Inform.* 2017;245:614.
10. Kim H-J, Barsevick AM, Tulman L, McDermott PA. Treatment-related symptom clusters in breast cancer: a secondary analysis. *J Pain Symptom Manage.* 2008;36(5):468-479.
11. Partridge AH, Burstein HJ, Winer EP. Side effects of chemotherapy and combined chemohormonal therapy in women with early-stage breast cancer. *J Natl Cancer Inst Monogr.* 2001;2001(30):135-142.
12. West M, Harrison J. *Bayesian Forecasting and Dynamic Models.* 2nd ed. New York, NY: Springer; 1997.
13. Carpenter B, Gelman A, Hoffman MD, et al. Stan: a probabilistic programming language. *J Stat Softw.* 2017;76(1):1-32.
14. R Core Team. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing; 2020.
15. Matsuura K. GitHub repository. https://github.com/MatsuuraKentarou/Matrix_Decomposition_in_Meta-Analysis_for_Adverse_Events; 2021.
16. Watanabe S. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J Mach Learn Res.* 2010;11:3571-3594.
17. Chang J, Gerrish S, Wang C, Boyd-graber J, Blei D. Reading tea leaves: how humans interpret topic models. *Advances in Neural Information Processing Systems 22*; New York, NY: Curran Associates Inc.; 2009:288-296.
18. Schofield A, Magnusson M, Mimno D. Pulling out the stops: rethinking stopword removal for topic models. Paper presented at: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers; 2017: 432-436.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Matsuura K, Tsuchida J, Ando S, Sozu T. Matrix decomposition in meta-analysis for extraction of adverse event pattern and patient-level safety profile. *Pharmaceutical Statistics.* 2021;20:806–819. <https://doi.org/10.1002/pst.2109>

APPENDIX A.

Case 1: one mechanism, multiple treatments, multiple studies, and multiple periods

The superscript m in $\mathbf{M}^{(m)}$ is removed. The log-probability Equation (15) is reduced to:

$$\begin{aligned}
 & \log P(\tilde{\boldsymbol{\theta}}^{(t)}, \mathbf{M}, \mathbf{R}^{(t)}, \tilde{\boldsymbol{\xi}}^{(s)}, \mathbf{S}^{(s)} \mid \mathbf{Y}_p, \sigma_M, \sigma_R, \sigma_S, \sigma_G) \\
 &= \sum_{p=1}^P \sum_{i=1}^{N_p} \sum_{j=1}^J \log \text{Poisson} \left(Y_{ij}^{(p)} \mid C_i^{(p)} \left[\exp \tilde{\boldsymbol{\theta}}^{(t_p)} \text{softmax}(\mathbf{M} + \mathbf{R}^{(t_p)}) + \exp \tilde{\boldsymbol{\xi}}^{(s_p)} \text{softmax}(\mathbf{S}^{(s_p)}) \right]_{ij} \right) \\
 &+ \sum_{k=1}^K \left[\sum_{j \in \mathfrak{J}} \log \mathcal{N}(M_{kj} \mid 0, \sigma_M^2) + \sum_{(j,j') \in \mathcal{J}} \log \mathcal{N}(M_{kj'} \mid M_{kj}, \sigma_G^2) \right] \\
 &+ \sum_{t=1}^T \sum_{k=1}^K \left[\sum_{j \in \mathfrak{J}} \log \mathcal{N}(R_{kj}^{(t)} \mid 0, \sigma_R^2) + \sum_{(j,j') \in \mathcal{J}} \log \mathcal{N}(R_{kj'}^{(t)} \mid R_{kj}^{(t)}, \sigma_G^2) \right] \\
 &+ \sum_{s=1}^S \sum_{l=1}^L \left[\sum_{j \in \mathfrak{J}} \log \mathcal{N}(S_{lj}^{(s)} \mid 0, \sigma_S^2) + \sum_{(j,j') \in \mathcal{J}} \log \mathcal{N}(S_{lj'}^{(s)} \mid S_{lj}^{(s)}, \sigma_G^2) \right]
 \end{aligned}$$

Case 2: one mechanism, one treatment, multiple studies, and multiple periods

In addition to the above case 1, the superscripts t in $\boldsymbol{\beta}^{(t)}$, $\boldsymbol{\theta}^{(t)}$, and $\boldsymbol{\phi}^{(t)}$ are removed. $\mathbf{R}^{(t)}$ and Equations (10) and (13) are eliminated. Equation (6) is simplified to: $\boldsymbol{\Phi}_k^{(t)} = \text{softmax}(\mathbf{M}_k)$. The log-probability Equation (15) is reduced to:

$$\begin{aligned} & \log P(\tilde{\boldsymbol{\theta}}, \mathbf{M}, \tilde{\boldsymbol{\xi}}^{(s)}, \mathbf{S}^{(s)} \mid \mathbf{Y}_p, \sigma_M, \sigma_S, \sigma_G) \\ &= \sum_{p=1}^P \sum_{i=1}^{N_p} \sum_{j=1}^J \log \text{Poisson} \left(Y_{ij}^{(p)} \mid C_i^{(p)} \left[\exp \tilde{\theta} \text{softmax}(\mathbf{M}) + \exp \tilde{\xi}^{(s_p)} \text{softmax}(\mathbf{S}^{(s_p)}) \right]_{ij} \right) \\ &+ \sum_{k=1}^K \left[\sum_{j \in \mathfrak{J}} \log \mathcal{N}(M_{kj} \mid 0, \sigma_M^2) + \sum_{(j,j') \in \mathcal{J}} \log \mathcal{N}(M_{kj'} \mid M_{kj}, \sigma_G^2) \right] \\ &+ \sum_{s=1}^S \sum_{l=1}^L \left[\sum_{j \in \mathfrak{J}} \log \mathcal{N}(S_{ij}^{(s)} \mid 0, \sigma_S^2) + \sum_{(j,j') \in \mathcal{J}} \log \mathcal{N}(S_{ij'}^{(s)} \mid S_{ij}^{(s)}, \sigma_G^2) \right] \end{aligned}$$

Case 3: one mechanism, one treatment, one study, and multiple periods

In addition to the above case 2, the superscripts s in $\boldsymbol{\alpha}^{(s)}$, $\boldsymbol{\xi}^{(s)}$, and $\boldsymbol{\eta}^{(s)}$ are removed. The log-probability Equation (15) is reduced to:

$$\begin{aligned} \log P(\tilde{\boldsymbol{\theta}}, \mathbf{M}, \tilde{\boldsymbol{\xi}}, \mathbf{S} \mid \mathbf{Y}_p, \sigma_M, \sigma_S, \sigma_G) &= \sum_{p=1}^P \sum_{i=1}^{N_p} \sum_{j=1}^J \log \text{Poisson} \left(Y_{ij}^{(p)} \mid C_i^{(p)} \left[\exp \tilde{\theta} \text{softmax}(\mathbf{M}) + \exp \tilde{\xi} \text{softmax}(\mathbf{S}) \right]_{ij} \right) \\ &+ \sum_{k=1}^K \left[\sum_{j \in \mathfrak{J}} \log \mathcal{N}(M_{kj} \mid 0, \sigma_M^2) + \sum_{(j,j') \in \mathcal{J}} \log \mathcal{N}(M_{kj'} \mid M_{kj}, \sigma_G^2) \right] \\ &+ \sum_{l=1}^L \left[\sum_{j \in \mathfrak{J}} \log \mathcal{N}(S_{ij} \mid 0, \sigma_S^2) + \sum_{(j,j') \in \mathcal{J}} \log \mathcal{N}(S_{ij'} \mid S_{ij}, \sigma_G^2) \right] \end{aligned}$$

Case 4: one mechanism, one treatment, one study, and one period (no baseline period)

In addition to the above case 3, $\boldsymbol{\alpha}^{(s)}$, $\boldsymbol{\xi}^{(s)}$, $\boldsymbol{\eta}^{(s)}$, and $\mathbf{S}^{(s)}$ are eliminated, and Equations (4), (7), (8), (11), and (14) are eliminated. The subscripts and superscripts p are removed. Equation (2) is simplified to $\lambda_{ij} = \beta_{ij}$, and the log-probability Equation (15) is reduced to:

$$\begin{aligned} \log P(\tilde{\boldsymbol{\theta}}, \mathbf{M} \mid \mathbf{Y}, \sigma_M, \sigma_G) &= \sum_{i=1}^N \sum_{j=1}^J \log \text{Poisson} \left(Y_{ij} \mid C_i \left[\exp \tilde{\theta} \text{softmax}(\mathbf{M}) \right]_{ij} \right) \\ &+ \sum_{k=1}^K \left[\sum_{j \in \mathfrak{J}} \log \mathcal{N}(M_{kj} \mid 0, \sigma_M^2) + \sum_{(j,j') \in \mathcal{J}} \log \mathcal{N}(M_{kj'} \mid M_{kj}, \sigma_G^2) \right] \end{aligned}$$