Matrix Decomposition for the Extraction of Adverse Event Patterns and Patient-Level Safety Profile for Each Drug in Combination Therapy

Kentaro Matsuura* Takashi Sozu†

Summary

The primary purpose of evaluating adverse events (AEs) in clinical studies is to understand the patterns of AEs caused by treatment at the population level. A secondary purpose is to know which AE patterns are likely to occur for each patient to support patient-level treatment strategies. To achieve these objectives simultaneously, an extended model of nonnegative matrix factorization (NMF) that can integrate data from multiple clinical studies has been proposed. An important challenge that has not yet been addressed is the estimation of AE patterns for each drug in combination therapy. In this report, we constructed a new statistical model to achieve this challenge. We illustrated the estimation of the AE patterns by applying our model to the data from two actual Phase III clinical studies of docetaxel and ramucirumab combination therapy.

Key words

Pattern extraction; Clinical trial; Co-occurrence;

1 Introduction

The assessment of adverse events (AEs) that occur during treatment is important for effective treatment. The primary goal of evaluating AEs in clinical studies is to understand the patterns of AEs caused by treatment at the population level. For this purpose, it is common to summarize the number of patients who experienced AEs for each treatment and report the frequency of occurrence (Garon et al., 2014; Crowe et al., 2009). However, it is difficult to use such population-level methods to determine patient-level treatment strategies. If we knew which AE patterns are likely to occur for each patient, which we call the patient-level safety profile in this report, it would be useful for treatment strategies. To simultaneously extract AE patterns and patient-level safety profiles, an extended model of nonnegative matrix factorization (NMF) (Lee and Seung, 1999) was proposed by Matsuura et al. (2021).

An example of the NMF for AEs and the main points of their model (Matsuura et al., 2021) are described below. Suppose we have the 4×5 occurrence matrix $\boldsymbol{Y}$ (Eq. (1)) representing the number of occurrences of five AEs (in order: constipation, stomatitis, weak nausea, strong nausea, and vomiting) for four patients

*Department of Management Science, Graduate School of Engineering, Tokyo University of Science
†Department of Information and Computer Technology, Faculty of Engineering, Tokyo University of Science

during the treatment period.

$$\underset{\boldsymbol{Y}}{\begin{pmatrix}1&1&1&1&0\\1&2&0&1&0\\0&0&1&2&0\\0&0&0&1&1\end{pmatrix}} \approx \underset{\boldsymbol{\theta}}{\begin{pmatrix}1.7&1.6\\2.1&0.5\\0.0&2.5\\0.0&1.8\end{pmatrix}} \underset{\boldsymbol{\phi}}{\begin{pmatrix}0.3&0.5&0.0&0.1&0.0\\0.0&0.0&0.3&0.6&0.1\end{pmatrix}} = \underset{\boldsymbol{\lambda}}{\begin{pmatrix}0.6&0.9&0.4&1.1&0.2\\0.8&1.1&0.1&0.5&0.1\\0.0&0.0&0.7&1.5&0.3\\0.0&0.0&0.5&1.1&0.2\end{pmatrix}} \tag{1}$$

For instance, the element in the first row and second column of $\boldsymbol{Y}$ represents a single occurrence of stomatitis in the first patient. Assuming that the number of treatment-specific AE patterns is limited, $\boldsymbol{Y}$ is approximated by the product of the rank-2 nonnegative matrices ($\boldsymbol{\theta}$ and $\boldsymbol{\phi}$) (NMF). Each row of $\boldsymbol{\phi}$ can be interpreted as an AE pattern common to all patients (row 1: constipation and stomatitis pattern, row 2: nausea and vomiting pattern), and each row of $\boldsymbol{\theta}$ can represent how much of these pattern components each patient has. The product of these matrices is $\boldsymbol{\lambda}$, which is a matrix that approximates $\boldsymbol{Y}$.

However, several issues must be solved when applying the NMF to actual data. Matsuura et al. (2021) devised the following: (i) It was assumed that $\boldsymbol{\lambda}$ can be expressed as the sum of the study-specific matrix and the treatment-specific matrix, and each matrix can be decomposed into the product of two matrices; (ii) the model was prepared to effectively utilize the severity of AE (specifically, in the above example, since 0.3 and 0.6 in nausea and vomiting patterns are both nausea and only differ in severity, Matsuura et al. used a probability distribution to constrain these values to be similar); (iii) because a single clinical study does not provide a sufficient amount of data to estimate the AE patterns, it is necessary to deal with data from multiple clinical studies simultaneously. To do this, Matsuura et al. assumed that the AE patterns were the same for the same treatment and that the AE patterns of drugs with the same mechanism of action were similar. A statistical model reflecting these three points was constructed and applied to actual clinical study data. The estimated AE patterns coincided with the medical background and demonstrated how to use the patient-level safety profile.

However, their model (Matsuura et al., 2021) represents the AE pattern of a combination therapy by summing up the effects of all drugs; thus, the AE pattern for each drug cannot be estimated. Therefore, the safety of a new drug cannot be properly assessed. This study aimed to develop a model to evaluate AE patterns separately for each drug in a combination therapy.

In Section 2, the statistical model and the estimation method are described. In Section 3, the results of the application of the method are demonstrated. In Section 4, we summarize and discuss our findings.

2 Analysis Method

2.1 Statistical Model

In this report, we consider a typical design of clinical studies using anticancer drugs. That is, we consider a study design in which the comparator arm involves treatment with standard drug A and the experimental arm uses combination therapy with A and experimental drug B. We are interested in the AE patterns caused by B.

We suppose that we have S clinical study datasets and define s as the index of each study. Each study has two arms: A comparator arm (treatment with standard drug A) and an experimental arm (combination therapy with A and experimental drug B). Each arm has two periods: The baseline period from registration to the start of treatment and the treatment period. To separate AEs according to treatment, we define p as the serial index of each period, as shown in Table 1. The total number of periods is denoted as P, and the number of patients in period p is denoted as N_p. The notations are summarized in Table 1.

Table 1: Data structure and indices.

Study index s	Arm	Period index p	Treatment	Number of patients	Occurrence matrix
1	comparator	1	baseline	N_1	$\boldsymbol{Y}_1$
1	comparator	2	A	N_2	$\boldsymbol{Y}_2$
1	experimental	3	baseline	N_3	$\boldsymbol{Y}_3$
1	experimental	4	A+B	N_4	$\boldsymbol{Y}_4$
2	comparator	5	baseline	N_5	$\boldsymbol{Y}_5$
2	comparator	6	A	N_6	$\boldsymbol{Y}_6$
2	experimental	7	baseline	N_7	$\boldsymbol{Y}_7$
2	experimental	8	A+B	N_8	$\boldsymbol{Y}_8$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
S	comparator	$P-3$	baseline	N_{P-3}	$\boldsymbol{Y}_{P-3}$
S	comparator	$P-2$	A	N_{P-2}	$\boldsymbol{Y}_{P-2}$
S	experimental	$P-1$	baseline	N_{P-1}	$\boldsymbol{Y}_{P-1}$
S	experimental	P	A+B	N_P	$\boldsymbol{Y}_P$

In this report, AEs with the same name but different severities are distinguished. The combination of the name and severity is identified as the "AE type". For example, "headaches" with severity 2 is "headaches_2." J is defined as the total number of AE types.

The number of occurrences is summarized per period p, and the result is expressed as $N_p \times J$ matrix $\boldsymbol{Y}_p$. $\boldsymbol{Y}_p$ is an occurrence matrix of AEs during the period p. $Y_{ij}^{(p)}$, which is the (i,j) element of $\boldsymbol{Y}_p$, represents the number of occurrences of AE type j for patient i during period p. In general, in clinical studies, AEs are recorded based on severity, start date, and end date. There are multiple ways to convert this form of data into the number of occurrences of AE types. In this report, the number of occurrences is first counted for each cycle, and finally, the sum of the counts for all cycles is defined as the number of occurrences of AE types. If an AE occurred persistently from the beginning to the end of a cycle, it is counted as 1.

We explain the statistical model for generating $\boldsymbol{Y}_p$ during period p in study s. Scalars are represented in italics, vectors in bold font, and matrices in bold italics. To simplify, $\boldsymbol{Y}_p$ is written as $\boldsymbol{Y}$ and N_p as N. The patient index is written as i and the AE type index as j. Because $\boldsymbol{Y}$ is a nonnegative integer data, each element of $\boldsymbol{Y}$ was assumed to be independently generated from a Poisson distribution as follows:

$$Y_{ij} \sim \text{Poisson}\left(C_i \lambda_{ij}\right) \quad i = 1, \ldots, N \quad j = 1, \ldots, J \tag{2}$$

where C_i is a constant representing the length of the treatment period of patient i. λ_{ij} is the (i,j) element of the $N \times J$ matrix $\boldsymbol{\lambda}$. $\boldsymbol{\lambda}$ is assumed to be separated into study-specific and treatment-specific terms. In the combination therapy of A and B, we assume that there is no interaction between A and B, and that the effects of A and B are additive:

$$\boldsymbol{\lambda} = \begin{cases} \boldsymbol{\alpha}^{(s)} & \text{if } p \text{ is a baseline period} \\ \boldsymbol{\alpha}^{(s)} + \boldsymbol{\beta}^{(A)} & \text{if } p \text{ is under treatment A} \\ \boldsymbol{\alpha}^{(s)} + \boldsymbol{\beta}^{(A)} + \boldsymbol{\beta}^{(B)} & \text{if } p \text{ is under treatment A+B} \end{cases} \tag{3}$$

$\boldsymbol{\alpha}^{(s)}$ is a study-specific parameter representing the occurrence intensity of each AE caused by disease, radiotherapy, etc. $\boldsymbol{\beta}^{(A)}$ and $\boldsymbol{\beta}^{(B)}$ are parameters per drug representing the occurrence intensity of each AE type caused by the drug. $\boldsymbol{\alpha}^{(s)}$ corresponds to the intercept term.

If the baseline periods are not included in the data, or if the number of occurrences of AEs in the baseline periods is very small, we can ignore the baseline periods, eliminate $\boldsymbol{\alpha}^{(s)}$, and use the following equation

instead of Eq. (3):

$$\boldsymbol{\lambda} = \begin{cases} \boldsymbol{\beta}^{(A)} & \text{if } p \text{ is under treatment A} \\ \boldsymbol{\beta}^{(A)} + \boldsymbol{\beta}^{(B)} & \text{if } p \text{ is under treatment A+B} \end{cases} \tag{4}$$

Because the occurrence patterns of AEs are usually limited, it is assumed that they consist of a combination of relatively few patterns with little information loss, and the matrix decomposition is applied as follows:

$$\boldsymbol{\beta}^{(d)} = \boldsymbol{\theta}^{(d)}\boldsymbol{\phi}^{(d)} \quad d = A, B \tag{5}$$

$$\boldsymbol{\alpha}^{(s)} = \boldsymbol{\xi}^{(s)}\boldsymbol{\eta}^{(s)} \tag{6}$$

where $\boldsymbol{\theta}^{(d)}$ is an $N \times K$ matrix, $\boldsymbol{\phi}^{(d)}$ is a $K \times J$ matrix, $\boldsymbol{\xi}^{(s)}$ is an $N \times L$ matrix, and $\boldsymbol{\eta}^{(s)}$ is an $L \times J$ matrix. K and L are $\ll J$ and represent the numbers of AE patterns. As shown in Eq. (1), it can be interpreted that $\boldsymbol{\theta}^{(d)}$ has different values for each patient, and $\boldsymbol{\phi}^{(d)}$ represents the AE patterns of drug d common to all patients. Similarly, it can be interpreted that $\boldsymbol{\xi}^{(s)}$ has different values for each patient, and $\boldsymbol{\eta}^{(s)}$ represents the AE patterns of study s common to all patients.

To estimate $\boldsymbol{\theta}^{(d)}$ as a nonnegative matrix, $\boldsymbol{\theta}^{(d)}$ is reparametrized as follows:

$$\boldsymbol{\theta}^{(d)} = \exp \widetilde{\boldsymbol{\theta}}^{(d)} \quad d = A, B \tag{7}$$

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

For $\boldsymbol{\phi}^{(d)}$, the existing knowledge of treatments is used, assuming that $\boldsymbol{\phi}^{(d)}$ is constructed as follows:

$$\boldsymbol{\phi}_k^{(d)} = \text{softmax}\left(\mathbf{M}_k^{(d)}\right) \quad d = A, B \quad k = 1, \ldots, K \tag{8}$$

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

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

This softmax function facilitates comparisons between patterns because the sum of the J elements of each row (pattern) of $\boldsymbol{\phi}^{(d)}$ becomes 1. This corresponds to the fact that the sum of each row of $\boldsymbol{\phi}$ is 1 in Eq. (1).

Similar constraints are assumed for $\boldsymbol{\alpha}^{(s)}$, as in $\boldsymbol{\beta}^{(d)}$. $\boldsymbol{\xi}^{(s)}$ is reparametrized as follows to make each element nonnegative:

$$\boldsymbol{\xi}^{(s)} = \exp \widetilde{\boldsymbol{\xi}}^{(s)} \tag{9}$$

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

$$\boldsymbol{\eta}_l^{(s)} = \text{softmax}\left(\mathbf{S}_l^{(s)}\right) \quad l = 1, \ldots, L \tag{10}$$

In an occurrence pattern, AE types with the same name and similar severities (e.g., headaches_2 and headaches_3) should have similar occurrence intensities. Therefore, $\boldsymbol{M}^{(d)}$ and $\boldsymbol{S}^{(s)}$ are constrained using a normal dynamic linear model (NDLM) (West and Harrison, 1997), which is often applied in dose-response curves as follows:

$$M_{kj'}^{(d)} \sim \mathcal{N}\left(M_{kj}^{(d)}, \sigma_G^2\right) \quad d = A, B \quad k = 1, \ldots, K \quad (j, j') \in \mathcal{J} \tag{11}$$

$$S_{lj'}^{(s)} \sim \mathcal{N}\left(S_{lj}^{(s)}, \sigma_G^2\right) \quad l = 1, \ldots, L \quad (j, j') \in \mathcal{J} \tag{12}$$

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. $\boldsymbol{M}^{(d)}$ and $\boldsymbol{S}^{(s)}$ are constrained as follows so that they can be identified and estimated:

$$M_{kj}^{(d)} \sim \mathcal{N}\left(0, \sigma_M^2\right) \quad d = A, B \quad k = 1, \ldots, K \quad j \in \mathfrak{J} \tag{13}$$

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

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

The main difference from their model (Matsuura et al., 2021) is shown in Eq. (3). For the A and B combination therapy, their model assumes that $\boldsymbol{\lambda} = \boldsymbol{\alpha}^{(s)} + \boldsymbol{\beta}^{(A+B)}$, and $\boldsymbol{\phi}^{(A)}$ and $\boldsymbol{\phi}^{(A+B)}$ are similar. Therefore, there is a disadvantage that $\boldsymbol{\phi}^{(B)}$ cannot be separated, and the experimental drug B cannot be evaluated.

2.2 Estimation

Considering Eqs. (2), (3), and (5) to (14) for all data $\boldsymbol{Y}_p$ $(p = 1, \ldots, P)$, the log of the posterior probability is constructed as follows:

$$\begin{aligned}
\log P = & \sum_{p=1}^{P} \sum_{i=1}^{N_p} \sum_{j=1}^{J} \log \mathrm{Poisson}\left(Y_{ij}^{(p)} \;\middle|\; C_i^{(p)} \lambda_{ij}^{(p)}\right) \\
& + \sum_{d \in \{A,B\}} \sum_{k=1}^{K} \left[\sum_{j \in \mathfrak{J}} \log \mathcal{N}\left(M_{kj}^{(d)} \;\middle|\; 0, \sigma_M^2\right) + \sum_{(j,j') \in \mathcal{J}} \log \mathcal{N}\left(M_{kj'}^{(d)} \;\middle|\; M_{kj}^{(d)}, \sigma_G^2\right) \right] \\
& + \sum_{s=1}^{S} \sum_{l=1}^{L} \left[\sum_{j \in \mathfrak{J}} \log \mathcal{N}\left(S_{lj}^{(s)} \;\middle|\; 0, \sigma_S^2\right) + \sum_{(j,j') \in \mathcal{J}} \log \mathcal{N}\left(S_{lj'}^{(s)} \;\middle|\; S_{lj}^{(s)}, \sigma_G^2\right) \right]
\end{aligned} \tag{15}$$

where the softmax function operates on each row of a matrix and returns a matrix, $\mathrm{Poisson}\left(y \mid \lambda\right)$ represents the probability mass at y of the Poisson distribution with parameter λ, and $\mathcal{N}\left(y \;\middle|\; \mu, \sigma^2\right)$ represents the probability density at y of the normal distribution with parameters μ and σ^2.

$\boldsymbol{\theta}^{(A)}$, $\boldsymbol{M}^{(A)}$, $\boldsymbol{\theta}^{(B)}$, $\boldsymbol{M}^{(B)}$, and $\boldsymbol{\xi}^{(s)}$, $\boldsymbol{S}^{(s)}(s = 1, \ldots, S)$ are estimated by maximizing the log-probability $\log P$ in Eq. (15). Bayes estimation can be performed by setting an appropriate prior distribution. σ_M, σ_S, and σ_G can also be estimated if there are sufficient data. In this report, they are regarded as hyperparameters and assigned fixed values. The number of patterns K and L are also hyperparameters that must be fixed before the estimation.

If the data do not have baseline periods, or if the number of occurrences of AEs in the baseline period is very small, $\boldsymbol{\theta}^{(A)}$, $\boldsymbol{M}^{(A)}$, $\boldsymbol{\theta}^{(B)}$, and $\boldsymbol{M}^{(B)}$ are estimated by maximizing the following log-probability, considering Eqs. (2), (4), (5), (7), (8), (11), and (13) as follows:

$$\begin{aligned}
\log P = & \sum_{p=1}^{P} \sum_{i=1}^{N_p} \sum_{j=1}^{J} \log \mathrm{Poisson}\left(Y_{ij}^{(p)} \;\middle|\; C_i^{(p)} \lambda_{ij}^{(p)}\right) \\
& + \sum_{d \in \{A,B\}} \sum_{k=1}^{K} \left[\sum_{j \in \mathfrak{J}} \log \mathcal{N}\left(M_{kj}^{(d)} \;\middle|\; 0, \sigma_M^2\right) + \sum_{(j,j') \in \mathcal{J}} \log \mathcal{N}\left(M_{kj'}^{(d)} \;\middle|\; M_{kj}^{(d)}, \sigma_G^2\right) \right]
\end{aligned} \tag{16}$$

2.3 Outputs

From the estimates of $\boldsymbol{M}^{(A)}$ and $\boldsymbol{M}^{(B)}$, the AE patterns of the standard drug A ($\boldsymbol{\phi}^{(A)}$) and the AE patterns of the experimental drug B ($\boldsymbol{\phi}^{(B)}$) can be obtained separately. The outputs of interest in this report are $\boldsymbol{\phi}^{(B)}$ and $\boldsymbol{\theta}^{(B)}$.

The estimate of $\boldsymbol{\phi}^{(B)}$ is a set of AE type patterns specific to treatment B. $\boldsymbol{\phi}_k^{(B)}$, the k-th row of $\boldsymbol{\phi}^{(B)}$ (a row vector), represents 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 their intensities in descending order and presenting them in a table. Patterns in which severe AEs are most likely to occur can also be confirmed.

The estimate of $\boldsymbol{\theta}^{(B)}$ represents the patient-level safety profiles of drug B. $\boldsymbol{\theta}_i^{(B)}$, the i-th row of $\boldsymbol{\theta}^{(B)}$ (a row vector), represents how much of these pattern components patient i has.

3 Application to Clinical Trial Data

In this section, we present an example of applying the analysis method described in Section 2 to actual clinical study data.

3.1 Data

We selected all data sets from Phase III studies involving docetaxel and ramucirumab combination therapy from Vivli (https://vivli.org), which contains patient-level AE information. As a result, two datasets (registration numbers: NCT01168973 and NCT00703326) were selected. The outlines of each arm included in the datasets are listed in Table 2.

Table 2: Data summary and indices.

Study index s	Arm	Period index p	Treatment	Number of patients	Total number of AEs
1	comparator	1	baseline	597	392
1	comparator	2	A	597	18237
1	experimental	3	baseline	617	424
1	experimental	4	A+B	617	24231
2	comparator	5	baseline	380	524
2	comparator	6	A	380	25536
2	experimental	7	baseline	749	1051
2	experimental	8	A+B	749	65861

Table 2 corresponds to the case where the number of studies in Table 1 is $S = 2$, docetaxel corresponds to treatment A, and docetaxel and ramucirumab correspond to treatment A + B. The datasets included AEs that occurred during the baseline or treatment periods. The two studies focused on different diseases, lung cancer and breast cancer. However, assuming that disease-specific AEs were observed in the baseline period, these studies were used simultaneously. The occurrence matrices of the AE types were created according to Section 2.1. The AE names used were the MedDRA High Level Terms (HLTs) recorded in the studies. Severity was defined according to the Common Terminology Criteria for Adverse Events (CTCAE; 1, mild; 2, moderate; 3, severe; 4, life-threatening; 5, death) and ranged from 1 to 5. The combination of name and severity was considered as the AE type. The total number of AE types was 1,225. Patients without AEs were excluded because their safety profiles were estimated to be zero, and their data did not contribute to the estimation of AE patterns. Consequently, 2,343 patients and 136,256 AEs were analyzed. The average number of AEs per patient was approximately 58, and the number of AEs in the baseline period was approximately 2% of the total number of AEs.

3.2 Analysis Method

Using the method described in Section 2.2, $\boldsymbol{\theta}^{(A)}$, $\boldsymbol{\phi}^{(A)}$, $\boldsymbol{\theta}^{(B)}$, $\boldsymbol{\phi}^{(B)}$, $\boldsymbol{\xi}^{(1)}$, $\boldsymbol{\eta}^{(1)}$, $\boldsymbol{\xi}^{(2)}$, and $\boldsymbol{\eta}^{(2)}$ were estimated via MAP estimation. The initial values were set to random numbers. The number of patterns and hyperparameters used were $K = 10$, $L = 1$, $\sigma_M = \sigma_S = 5$, and $\sigma_G = 1.5$. These values were determined from experience, but there is room for consideration (Matsuura et al., 2021). C_i was defined as the number of treatment days/84. The software used for this estimation was TensorFlow 1.3 (Abadi, 2016). MAP estimation was performed using the Adam algorithm (Kingma, 2015).

3.3 Results

This section describes the outputs mentioned in Section 2.3. Here, the $\boldsymbol{\phi}^{(B)}$ patterns of treatment with ramucirumab are shown. Ten patterns of $\boldsymbol{\phi}^{(B)}$ were numbered in descending order of the log-probability increase. The results for the top five patterns with a large log-probability increase are shown in Table 3. The other patterns are shown in Appendix.

Table 3: Result for $\boldsymbol{\phi}^{(B)}$ (pattern 1 to 5).

Pattern k	AE type	$\phi_{kj}^{(B)}$
1	cardiovascular neoplasms benign_1	9.80
	pneumothorax and pleural effusions nec_2	7.25
	oedema nec_3	5.75
	elevated cholesterol_2	5.38
	haemorrhoids and gastrointestinal varices (excl oesophageal)_2	5.26
	liver function analyses_1	4.87
	upper respiratory tract signs and symptoms_2	4.45
	tissue enzyme analyses nec_2	3.88
	death and sudden death_5	**0.03**
	oncologic complications and emergencies_5	**0.02**
2	nasal disorders nec_1	53.16
	vascular hypertensive disorders nec_2	19.98
	vascular hypertensive disorders nec_1	10.16
	upper respiratory tract infections_1	7.94
	telangiectasia and related conditions_1	1.50
	upper respiratory tract infections_2	1.49
	dermatitis and eczema_1	1.14
	emotional and mood disturbances nec_1	0.76
	upper respiratory tract infections_5	**0.01**
	death and sudden death_5	**0.00**
3	nail and nail bed conditions (excl infections and infestations)_2	33.91
	gingival haemorrhages_1	13.87
	peripheral neuropathies nec_3	5.32
	gastrointestinal and abdominal pains (excl oral and throat)_1	4.80
	skin and subcutaneous conditions nec_3	4.55
	lid, lash and lacrimal infections, irritations and inflammations_1	3.08
	total fluid volume increased_1	3.03
	pruritus nec_1	2.90
	aortic aneurysms and dissections_5	**0.00**
	breast and nipple neoplasms malignant_5	**0.00**
4	vascular hypertensive disorders nec_3	23.91
	disturbances in initiating and maintaining sleep_2	8.76
	ocular disorders nec_1	7.78
	oral dryness and saliva altered_1	6.57
	thrombocytopenias_1	5.40
	hyperpigmentation disorders_1	4.32
	thrombocytopenias_2	3.84
	musculoskeletal and connective tissue pain and discomfort_2	3.61
	sepsis, bacteraemia, viraemia and fungaemia nec_5	**0.03**
	non-site specific procedural complications_5	**0.02**
5	sensory abnormalities nec_1	19.35
	appetite disorders_1	15.18
	diarrhoea (excl infective)_1	15.17
	anxiety symptoms_1	11.87
	joint related signs and symptoms_2	11.66
	skin and subcutaneous conditions nec_2	4.66
	dyspeptic signs and symptoms_2	3.25
	dyspeptic signs and symptoms_1	2.43
	bacterial infections nec_5	**0.01**
	pulmonary oedemas_5	**0.01**

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 are presented for each pattern. We also show the top two AE types with severity 5 (bold letters).

We focused on patterns 1 and 4, which had high values of $\phi_{kj}^{(B)}$ for death (severity 5) among the top five patterns. Pattern 1 was interpreted as a "neoplasm pattern", in which cardiovascular neoplasms benign of severity 1, oedema not elsewhere classified (NEC) of severity 3, haemorrhoids and gastrointestinal varices of severity 2, and death by oncologic emergency tended to occur simultaneously. Pattern 4 was a "severe vascular hypertensive disorders and infection pattern" involving vascular hypertensive disorders NEC of severity 3, ocular disorders NEC, oral dryness and saliva altered, thrombocytopenias, and death by sepsis, bacteraemia, viraemia and fungaemia.

As an example of a patient-level safety profile, $\boldsymbol{\theta}_{i'}^{(B)}$ for patient i' is shown in Figure 1. This patient had a high value of the pattern 4 component. Because thrombocytopenias of severity 1 to 3 occurred in this patient and these AE types are pattern 4 specific, $\theta_{i'4}^{(B)}$ was high.

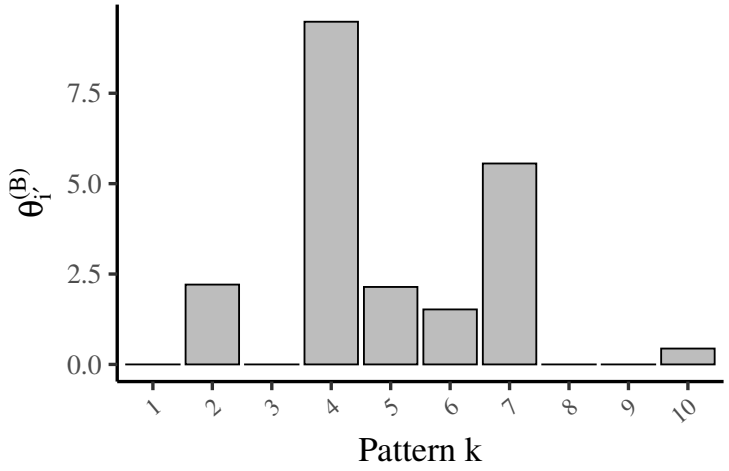

Figure 1: Estimated $\boldsymbol{\theta}_{i'}^{(B)}$.

We confirmed the responsiveness of the 1,366 patients administered docetaxel and ramucirumab. For patterns 1 and 4, the distributions of the elements of $\boldsymbol{\theta}_{:k}^{(B)}$ (the column k of $\boldsymbol{\theta}^{(B)}$) are shown in Figure 2. The distributions of $\boldsymbol{\theta}_{:k}^{(B)}$ were similar to the zero-inflated lognormal distribution.

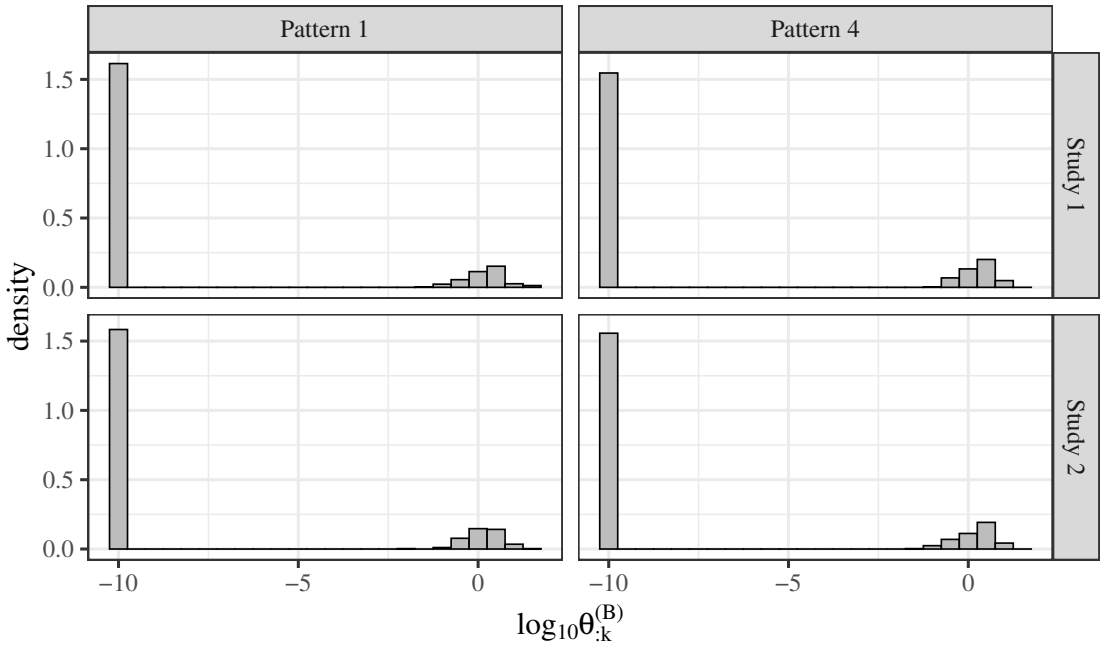

Figure 2: Distributions of $\theta_{:k}^{(B)}$ where values less than 10^{-10} were converted to 10^{-10} before taking the logarithm.

4 Discussion

A new model was constructed to estimate AE patterns separately for each drug in combination therapy by extending the model based on matrix factorization for patient-level AE data (Matsuura et al., 2021).

Although the combination therapy of the two drugs is described, our model can be applied to the combination therapy of three or more drugs. We assume that there is no interaction between drugs. If the interaction term should be considered in the model, $\boldsymbol{\lambda} = \boldsymbol{\alpha}^{(s)} + \boldsymbol{\beta}^{(A)} + \boldsymbol{\beta}^{(B)} + \boldsymbol{\beta}^{(A)} \odot \boldsymbol{\beta}^{(B)}$ (where $\odot$ represents the elementwise product) can be used in Eq. (3). Other methods may also be considered.

The parameters were estimated by assigning true values for the number of patterns and hyperparameter values. However, in practice, there is no way of knowing them. A nonparametric Bayesian method is one of the methods that enables the estimation of the number of patterns. Further studies and discussions of application to actual data are required to determine whether a nonparametric Bayesian method and estimation of hyperparameters are effective in our model.

The patient-level safety profile provides enriched information on the number of AE occurrences through AE patterns. If this safety profile can be predicted by using patient information including age, sex, weight, baseline laboratory observations, and treatment history, it would be useful because it would enable us to understand the risk before starting treatment. As shown in Section 3.3, the estimated distribution of $\boldsymbol{\theta}$ is often bimodal. Therefore, we can convert the elements of $\boldsymbol{\theta}$ to binary values of non-responders or responders by setting thresholds and use them as the response variables, use patient information as the predictors, and use logistic regression or a decision tree as the prediction model. In contrast, there may be ways to incorporate patient information directly into the statistical model before estimating $\boldsymbol{\theta}$. The effective use of patient information will be a subject for future work.

Acknowledgements

This report is based on research using data from data contributors Eli Lilly and Company that has been made available through Vivli, Inc. Vivli has not contributed to or approved, and is not in any way responsible for, the contents of this publication.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al. (2016). Tensorflow: a system for large-scale machine learning. *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*. 265–283. doi: 10.5555/3026877.3026899.

Crowe, B. J., Xia, H. A., Berlin, J. A., Watson, D. J., Shi, H., Lin, S. L., Kuebler, J., Schriver, R. C., Santanello, N. C., Rochester, G., et al. (2009). Recommendations for safety planning, data collection, evaluation and reporting during drug, biologics and vaccine development: a report of the safety planning, evaluation, and reporting team. *Clin. Trials*, **6**, 430–440. doi: 10.1177/1740774509344101.

Garon, E. B., Ciuleanu, T. E., Arrieta, O., Prabhash, K., Syrigos, K. N., Goksel, T., Park, K., Gorbunova, V., Kowalyszyn, R. D., Pikiel, J., et al. (2014). Ramucirumab plus docetaxel versus placebo plus docetaxel for second-line treatment of stage IV non-small-cell lung cancer after disease progression on platinum-based therapy (REVEL): a multicentre, double-blind, randomised phase 3 trial. *Lancet*, **384**, 665–673. doi: 10.1016/S0140-6736(14)60845-X.

Kingma, D. P. and Ba, J. (2015). Adam: A method for stochastic optimization. *Proceedings of the 3rd International Conference of Learning and Representation*. dblp: journals/corr/KingmaB14.

Lee, D. D. and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, **401**, 788–791. doi: 10.1038/44565.

Matsuura, K., Tsuchida, J., Ando, S., Sozu, T. (2021). Matrix decomposition in meta-analysis for extraction of adverse event pattern and patient-level safety profile. *Pharmaceut Stat.*, **20**, 806–819. doi: 10.1002/pst.2109.

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models, 2nd ed.* NY: Springer. doi:10.1007/b98971.

Appendix

The other patterns of $\boldsymbol{\phi}^{(B)}$ are shown in Table 4.

Table 4: Result for $\boldsymbol{\phi}^{(B)}$ (pattern 6 to 10).

Pattern k	AE type	$\phi_{kj}^{(B)}$
6	lacrimation disorders_1	50.88
	oedema nec_1	47.20
	cartilage disorders_2	0.27
	panniculitides_2	0.25
	lower limb fractures and dislocations_2	0.20
	oedema nec_2	0.12
	conjunctival and corneal bleeding and vascular disorders_2	0.12
	lacrimation disorders_2	0.11
	lower respiratory tract and lung infections_5	**0.00**
	death and sudden death_5	**0.00**
7	asthenic conditions_3	26.96
	lacrimation disorders_2	11.60
	gastrointestinal atonic and hypomotility disorders nec_2	7.01
	nasal disorders nec_2	3.92
	nausea and vomiting symptoms_1	3.91
	urinary abnormalities_2	3.48
	urinary abnormalities_1	3.39
	magnesium metabolism disorders_1	3.04
	renal failure and impairment_5	**0.03**
	sepsis, bacteraemia, viraemia and fungaemia nec_5	**0.03**
8	asthenic conditions_2	54.46
	appetite disorders_2	9.95
	paraesthesias and dysaesthesias_1	5.07
	dermal and epidermal conditions nec_2	3.61
	peritoneal and retroperitoneal disorders_1	2.89
	physical examination procedures and organ system status_3	2.14
	nausea and vomiting symptoms_2	2.05
	elevated cholesterol_1	0.95
	metastases to specified sites_5	**0.04**
	renal failure and impairment_5	**0.04**
9	joint related signs and symptoms_1	22.31
	stomatitis and ulceration_2	14.63
	headaches nec_1	11.87
	stomatitis and ulceration_1	5.93
	erythemas_1	3.38
	upper respiratory tract infections_2	2.77
	neutropenias_4	2.61
	febrile disorders_1	2.41
	neoplasms unspecified malignancy and site unspecified nec_5	**0.06**
	neutropenias_5	**0.05**
10	upper respiratory tract signs and symptoms_1	28.77
	pneumothorax and pleural effusions nec_1	14.93
	stomatitis and ulceration_1	13.47
	skin and subcutaneous conditions nec_1	6.50
	nail and nail bed conditions (excl infections and infestations)_3	5.66
	musculoskeletal and connective tissue pain and discomfort_1	3.86
	skin neoplasms benign_1	2.67
	oral soft tissue pain and paraesthesia_1	2.21
	general signs and symptoms nec_5	**0.01**
	respiratory tract disorders nec_5	**0.01**

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 are presented for each pattern. We also show the top two AE types with severity 5 (bold letters).