

## Research Article

# An Optimal Algorithm for Determining Risk Factors for Complex Diseases: Depressive Disorder, Osteoporosis, and Fracture in Young Patients with Breast Cancer Receiving Curative Surgery

Chieh-Yu Liu <sup>1,2</sup> and Chun-Hung Chang<sup>3</sup>

<sup>1</sup>Biostatistical Consulting Lab, Department of Speech Language Pathology and Audiology, National Taipei University of Nursing and Health Sciences, Taipei, Taiwan

<sup>2</sup>Department of Midwifery and Women Health Care, School of Nursing, National Taipei University of Nursing and Health Sciences, Taipei, Taiwan

<sup>3</sup>Department of Psychiatry & Brain Disease Research Center, China Medical University Hospital, Taichung, Taiwan

Correspondence should be addressed to Chieh-Yu Liu; [chiehyu@ntunhs.edu.tw](mailto:chiehyu@ntunhs.edu.tw)

Received 7 February 2018; Accepted 14 May 2018; Published 4 July 2018

Academic Editor: Panayiotis Vlamos

Copyright © 2018 Chieh-Yu Liu and Chun-Hung Chang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study proposed a novel algorithm to investigate the risk factors for complex diseases. We employed the novel algorithm to determine the risk factors for depressive disorder, osteoporosis, and fracture in young patients with breast cancer who were receiving curative surgery. The novel algorithm has three steps. First, multiple correspondence analysis (MCA) is used to transform the raw data set into a multidimensional coordinate matrix. Second, the expectation-maximization (EM) algorithm is used for clustering the multidimensional coordinates for each category of variable. Third,  $v$ -fold cross-validation is incorporated into the coordinate matrix obtained using the MCA-EM algorithm to determine the optimal clustering of complex diseases and risk factors. A total of 4108 patients with breast cancer aged 20–39 years were enrolled. The results revealed that depressive disorder, osteoporosis, and fracture were clustered with liver cirrhosis, chronic obstructive pulmonary disease (COPD), distant metastasis, and primary metastatic and adjuvant therapies, namely, chemotherapy, radiotherapy, tamoxifen, aromatase inhibitors, and trastuzumab. Among the risk factors identified using this novel algorithm, liver cirrhosis and COPD have been rarely mentioned in the literature. In conclusion, the novel algorithm proposed in this study enables physicians and clinicians to identify risk factors for multiple diseases.

## 1. Introduction

Patients with cancer experience numerous side effects or complications as a result of chemotherapy, radiotherapy, surgery, or other treatments [1, 2]. Due to recent advances in new treatments, particularly for patients with breast cancer, the survival duration of patients has been significantly prolonged [3, 4]. However, breast cancer remains the most prevalent malignant neoplasm diagnosed among women worldwide and is the leading cause of death among female cancers [5–7]. In Taiwan, breast cancer is the most prevalent malignant neoplasm [8]. Aging is a high risk factor for breast cancer [9, 10], and menopause may also be considered a

crucial risk factor for breast cancer [9, 11]. However, younger patients with breast cancer can receive (or tolerate) more treatments than older patients, which implies that younger patients are likely to experience more complications or side effects, such as fatigue, vomiting, and hair loss, which are common among women with breast cancer receiving chemotherapy or radiotherapy. In addition, some patients may exhibit more severe complications (or diseases), such as depressive disorder [12], osteoporosis [13], and fracture [14]. Therefore, patients with cancer may experience multiple side effects, complications, or diseases simultaneously. However, most published studies have investigated the risk factors associated with a single disease [12, 14, 15]; few have

investigated the risk factors associated with multiple diseases resulting from adjuvant treatments of breast cancer. This study developed a novel algorithm to determine the risk factors for multiple diseases, namely, depressive disorder, osteoporosis, and fracture, in young patients with breast cancer receiving curative surgery. Data in the Taiwan National Health Insurance Research Database (NHIRD) were employed by the algorithm.

## 2. Materials and Methods

**2.1. Study Database.** The study data were retrieved from a population-based claims database, the NHIRD, in Taiwan. The National Health Insurance (NHI) program in Taiwan was launched on March 1, 1995, and it provided health care coverage to more than 99% of the Taiwanese population in 2010 [16]. The NHIRD contains population-based health care information, including outpatient and inpatient clinic or hospital visits, dental service visits, and traditional Chinese medicine services. The diagnostic and medical procedures for diseases are based on the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) and Procedure Coding System for every medical service claim.

**2.2. Ethics Statement.** The ethical review of the present study was approved by the Institutional Review Board of the School of Nursing, National Taipei University of Nursing and Health Sciences (CN-IRB-2011-063). In the NHIRD, the personal information of study patients is fully encrypted using double encryption procedures by the National Health Insurance Administration (NHIA). Because this study had a secondary database study design, written informed consent forms from the enrolled patients did not need to be obtained. The NHIA fully guarantees the confidentiality of the personal and health information of the study patients.

**2.3. Study Population Selection Process.** At the beginning, we retrieved from NHIRD from 2001 to 2007 by using the breast cancer ICD-9-CM code = 174.XX, and 73,385 patients were selected. Because we need to select newly diagnosed breast cancer patients, we need to exclude breast cancer patients who were diagnosed before January 1, 2002 ( $n = 34,533$ ). Besides, we excluded death patients during 2001 to 2007 ( $n = 1353$ ). We also excluded patients diagnosed with baseline depressive disorder (ICD-9-CM codes: 296.2X-296.3X, 300.4, and 311.X), bipolar disorder (ICD-9-CM codes: 296.0, 296.1, 296.4, 296.5, 296.6, 296.7, 296.8, 296.80, and 296.89), alcohol-use-related mental disorders (ICD-9-CM codes: V113, 9800, 2650, 2651, 3575, 4255, 3050, 291, 303, and 571.0-571.3), and osteoporosis (ICD-9-CM code: 733.XX) or with fracture history (ICD-9-CM code: 800.XX-829.XX), and male breast cancer patients ( $n = 18$ ). There were 32,776 newly diagnosed breast cancer patients left. Because this study was aimed to investigate risk factors for complex diseases, including depressive disorder, osteoporosis, and fracture in young breast cancer patients receiving curative surgery, patients aged 20-39 years who had received a new diagnosis of breast cancer (ICD-9-CM code 174.XX)

between January 1, 2001, and December 31, 2007, were further selected. All young patients with breast cancer who received curative surgery for the first time from 2001 to 2007 were finally recruited. Young patients with breast cancer who developed depressive disorder (ICD-9-CM codes: 296.2X-296.3X, 300.4, and 311.X), osteoporosis (ICD-9-CM code: 733.XX), and fracture (ICD-9-CM codes: 800.XX-829.XX) after curative surgery were enrolled. At the end of the selection process, a total number of 4108 young breast cancer patients were selected for this study. The patient recruitment scheme is presented in Figure 1.

**2.4. Algorithm.** The research database contains several categorical variables including binomial or multinomial variables. The first step of data analysis is to convert a data matrix containing categorical variables into a matrix containing index variables (0 or 1) through multiple correspondence analysis (MCA) [17]; the resultant matrix is called the Burt table, and each index variable indicates each level in all categorical variables. Furthermore, each index variable can be transformed into Euclidean coordinates in a higher dimensional space. Second, the resulting Euclidean coordinate matrix obtained through MCA can be considered as a high-dimensional data set or mixture distribution data set; the EM algorithm has been proven effective in determining hidden clusters among a high-dimensional or mixture distribution data set [18-20]. Third, after determining the optimal number of clusters, we used the  $\nu$ -fold cross-validation technique to identify the optimal clustering. The steps of the whole algorithm are detailed as follows:

*Step 1 (MCA).* Let  $\mathbf{M}_{I \times K}$  be the raw data matrix with  $I$  as the subjects and  $k$  as the categorical variables.

- (1) Convert the raw data matrix into the so-called Burt matrix:
  - (i) If one categorical variable is the binary variable, let it be considered the original variable type in the Burt matrix.
  - (ii) If one categorical variable has more than two levels (i.e.,  $J_k > 2$  levels), convert this variable into a so-called indicator matrix,  $\mathbf{I} \times J_k$ , in which each column contains binary variables coded 0 or 1.
  - (iii) Place all the binary variable columns together to form the indicator matrix  $\mathbf{X}_{I \times J}$ .
  - (iv) Generate the Burt matrix as  $= \mathbf{X}'\mathbf{X}$ .
- (2) Calculate the column and row coordinates:
  - (i) The grand total of  $\mathbf{M}_{I \times K}$  is  $\mathbf{N}$ ; calculate the probability matrix as  $\mathbf{P} = \mathbf{N}^{-1}\mathbf{X}$ .
  - (ii) Let  $\mathbf{r}$  denote the vector of the row totals of  $\mathbf{P}$  (i.e.,  $\mathbf{r} = \mathbf{P}\mathbf{1}$ , where  $\mathbf{1}$  is a conformable vector comprised of 1's), and  $\mathbf{c}$  denote the vector of the

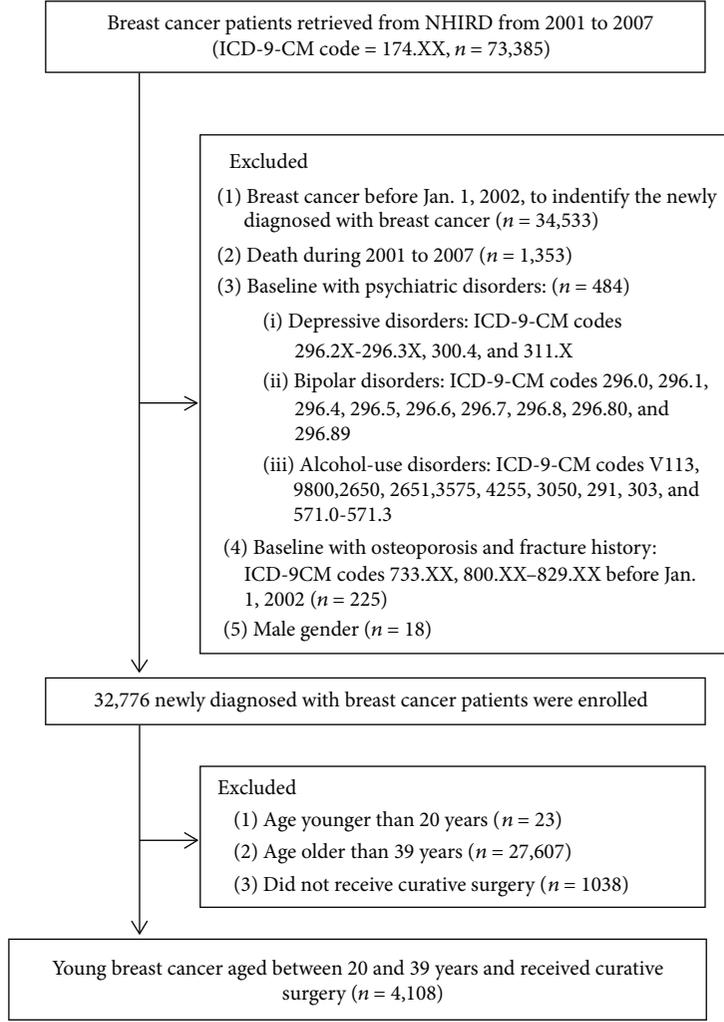


FIGURE 1: Recruitment scheme outlining patient selection.

column totals of  $\mathbf{P}$ ;  $\mathbf{Dc} = \text{diag}\{\mathbf{c}\}$ , and  $\mathbf{Dr} = \text{diag}\{\mathbf{r}\}$ .

- (iii) Calculate the coordinate scores by using the singular value decomposition method:

$$D_r^{-1/2}(Z - \mathbf{rc}^T)D_c^{-1/2} = P\Delta Q^T, \quad (1)$$

where  $\Delta$  is the diagonal matrix of singular values, and  $\Lambda = \Delta^2$  is the matrix containing eigenvalues. The so-called row coordinates and column coordinates are thus obtained as follows:

$$\begin{aligned} \text{row coordinate matrix} &= \mathbf{F} = D_r^{-1/2}P\Delta, \\ \text{column coordinate matrix} &= \mathbf{G} = D_c^{-1/2}Q\Delta. \end{aligned} \quad (2)$$

- (3) Determine the number of dimensions by using inertia estimation:

- (i) The Pearson chi-squared ( $\chi^2$ ) based distances from rows and columns to their respective point coordinate centers are calculated as

$$d_r = \text{diag}\{\mathbf{FF}^T\} \text{ and } d_c = \text{diag}\{\mathbf{GG}^T\}. \quad (3)$$

- (ii) If we select a subset of  $\mathbf{F}$  or  $\mathbf{G}$ , the *inertia* for the row coordinates and column coordinates for each level is given by

$$\begin{aligned} \text{Inertia}_r &= \frac{\text{diag}\{\mathbf{F}'\mathbf{F}'^T\}}{N}, \\ \text{Inertia}_c &= \frac{\text{diag}\{\mathbf{G}'\mathbf{G}'^T\}}{N}, \end{aligned} \quad (4)$$

where  $\mathbf{F}'$  and  $\mathbf{G}'$  are the subsets of  $\mathbf{F}$  and  $\mathbf{G}$ .

*Step 2.* Expectation-maximization (EM) algorithm for clustering

- (1) For a high-dimensional data set or mixture distribution data set, assuming that each variable in  $\mathbf{F}$  or  $\mathbf{G}$  is a random variable, the matrix can be modeled as a mixture normal-distribution probability density function as

$$\mathbf{A}_{\mu_i, \sigma_i}(x) = \frac{1}{\sigma_i \sqrt{2\pi}} e^{-(x-\mu_i)^2/2\sigma_i^2}, \quad (5)$$

where  $\mu_i$  and  $\sigma_i$  are the mean and standard deviation for each variable in  $\mathbf{F}$  or  $\mathbf{A}$ .

- (2) We use coordinates for each level of each categorical variable obtained using the MCA method, assuming that the number of dimensions selected using *Inertia*, which is previously obtained through MCA, is  $m$ .

- (i) Combine  $m$  density functions to model a mixture distribution:

$$P_\theta(x) = \sum_{i=1}^m \alpha_i P_{\theta_i}(x), \quad (6)$$

$$\theta = (\alpha_1, \dots, \alpha_m, \theta_1, \dots, \theta_m).$$

The  $K$ th normal distribution is parametrized using

$$\theta_k = \{\mu_k, \sigma_k\}. \quad (7)$$

- (ii) Given  $\theta$ , the log-likelihood function of  $x$  is as follows:

$$\log P_\theta(x) = \log \left( \sum_{i=1}^m \alpha_i P_{\theta_i}(x) \right). \quad (8)$$

- (3) For EM formulation, maximize the following function if  $\theta'$  exists which can maximize the following:

$$Q(\theta, \theta') = \sum_y P_{\theta'_y}(y | x) \log(\alpha_y P_{\theta'_y}(x)). \quad (9)$$

- (i) E-step (expectation step): compute the likelihood of  $y_i = k$ .

$$\Lambda_{i,k} = P_{\theta'_k}(y_i = k | x_i) = \frac{\alpha'_k P_{\theta'_k}(x_i)}{P_{\theta'_k}(x_i)} = \frac{\alpha'_k P_{\theta'_k}(x_i)}{\sum_{j=1}^m \alpha'_j P_{\theta'_j}(x_i)}. \quad (10)$$

- (ii) M-step (maximization step): update  $\alpha_k$ ,  $\mu_k$ , and  $\sigma_k$  for each categorical variable; for example,  $k = 1, \dots, m$ :

$$\alpha_k = \frac{1}{n} \sum_{i=1}^n \Lambda_{i,k},$$

$$\mu_k = \frac{\sum_{i=1}^n x_i \Lambda_{i,k}}{\sum_{i=1}^n \Lambda_{i,k}}, \quad (11)$$

$$\sigma_k^2 = \frac{\sum_{i=1}^n \Lambda_{i,k} \|x_i - \mu_k\|^2}{\sum_{i=1}^n \Lambda_{i,k}}.$$

*Step 3.* Apply  $\nu$ -fold cross-validation to the results of Step 2 to determine the optimal number of clusters

- (i) Randomly divide the overall sample into a number of  $\nu$ -folds (in this study, we used  $\nu = 10$ ).
- (ii) All the levels of categorical variables with coordinates can be classified into  $(\nu - 1)$ -folds as the training sample and 1 testing sample, which can be applied to  $i = 1, \dots, \nu$ th-fold.
- (iii) The results for the  $\nu$  replications are aggregated using  $-2^* \log(\text{likelihood})$  as the measurement of clustering cost (the lower, the more favorable) and are shown in a scree plot for determining the optimal number of clusters.

The MY Structured Query Language was used for database preprocessing, which comprised extraction, linkage, and cleaning of NHIRD data in this study. After the database was processed based on the inclusion and exclusion criteria, the study data sets were obtained. All statistical analyses were performed using STATISTICA (version 10 for Windows; Statistica, Tulsa, OK, USA), and a two-tailed  $p < 0.05$  was considered statistically significant.

### 3. Results

According to the recruitment scheme outlining patient selection (Figure 1), 4108 patients aged between 20 and 39 years who had received a diagnosis of breast cancer between January 1, 2001, and December 31, 2007, were recruited. The mean age of the patients was 34.6 years (standard deviation [SD] = 3.7 years). Of all the patients, 3.1% had depressive disorder, 0.7% had a fracture event, and 1.7% had osteoporosis. Regarding comorbidities, 1.7% patients had diabetes mellitus, 1.8% had hypertension, 0.7% had a history of heart failure, 0.5% had coronary heart disease, 0.4% had cerebrovascular disease, 2.3% had autoimmune disease, 0.9% had kidney disease, 0.4% had renal disease, 0.3% had liver cirrhosis, and 5.3% had chronic obstructive pulmonary disease (COPD). In addition, 13.4% exhibited distant metastasis and 6.9% exhibited primary metastasis. Regarding adjuvant therapies, 52.7% were receiving chemotherapy, 23.9% were receiving radiotherapy, 63.9% were receiving tamoxifen-related treatments, 7.5% were receiving aromatase inhibitor- (AI-) related treatments, and 3.2% were receiving trastuzumab treatments (Table 1).

After incorporating  $\nu$ -fold cross-validation into the EM clustering algorithm of the MCA coordinate matrix, four

TABLE 1: Demographic information of study sample ( $n = 4108$ ).

Variable	Mean (SD)	$n$	%
Age (yrs)	34.6 (3.7)		
CCI	1.9 (3.0)		
Comorbidities:			
Diabetes mellitus (DM)		69	1.7
Hypertension		72	1.8
Heart failure		27	.7
Coronary heart diseases (CHD)		20	.5
Cerebrovascular diseases		18	.4
Autoimmune diseases		95	2.3
Kidney diseases		36	.9
Renal diseases		17	.4
Liver cirrhosis		12	.3
COPD		216	5.3
Distant metastatic		550	13.4
Primary metastatic		284	6.9
Adjuvant therapies:			
Chemotherapy		2165	52.7
Radiotherapy		983	23.9
Tamoxifen		2623	63.9
AIs		309	7.5
Trastuzumab		132	3.2
Outcome variables:			
Depression disorders		126	3.1
Fracture		31	.7
Osteoporosis		71	1.7

CCI: Charlson comorbidity index; COPD: chronic obstructive pulmonary disease; AIs: aromatase inhibitors.

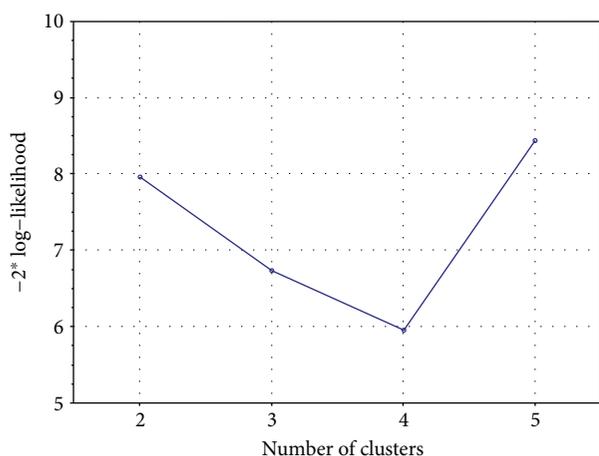


FIGURE 2: Graph of clustering cost sequence.

clusters were revealed to have the smallest  $-2 * \log\text{-likelihood}$  value; therefore, a four-cluster classification was defined as the optimal clustering approach (Figure 2).

Plots of each level of all variables in this study are presented in Figures 3(a) and 3(b). Figure 3(a) presents the four clusters obtained for all the points (levels of categorical

variables in the study) in four colors without labels in a three-dimensional scatter plot, whereas Figure 3(b) has each point with labels. The three outcome variables—depressive disorder, fracture, and osteoporosis—were clustered with chemotherapy, radiotherapy, tamoxifen, AIs, trastuzumab, primary metastasis, distant metastasis, and comorbidities including liver cirrhosis and COPD (Figure 3(b)). The clustering results are also tabulated in the right column of Table 2, together with a comparison with two published studies.

#### 4. Discussion

The present study proposed a novel algorithm and used the NHIRD, a population-based database, to investigate risk factors for multiple diseases, namely, depressive disorder, osteoporosis, and fracture, among young patients with breast cancer who were receiving curative surgery. The study results revealed that depressive disorder, osteoporosis, and fracture were clustered with chemotherapy, radiotherapy, tamoxifen, AIs, and trastuzumab treatments; this finding is in agreement with the findings of other studies [12, 14]. In addition, depressive disorder, osteoporosis, and fracture were clustered with distant metastasis and primary metastasis. Depressive disorder was associated with distant and primary metastases; these associations were proposed in published studies [21, 22]. However, depressive disorder, osteoporosis, and fracture were also clustered with liver cirrhosis and COPD, which have rarely been investigated. Liver cirrhosis has been linked to increased levels of estrogen, which may be a risk factor for breast cancer [23]. In addition, a study from 1992 demonstrated that breast cancer may be associated with antigen CA-153 in liver cirrhosis [24]; however, more recently published articles have rarely mentioned this association. Studies on the association of COPD with depressive disorder, osteoporosis, and fracture among young patients with breast cancer have not been conducted in recent years; therefore, further investigation of the association of COPD with depressive disorder, osteoporosis, and fracture among young patients with breast cancer may be required.

The present study adopted a different approach to identifying risk factors for multiple diseases: we used “clustering” instead of typical statistical analysis methods (e.g., logistic regression and Cox regression). However, most studies have addressed univariate outcome variables (e.g., overall survival for death outcome, disease- (or progression-) free survival for recurrence of some disease, and binary outcome variable with/without some disease). Studies that have employed univariate statistical analysis methods can provide information on risk factors for some specific disease onset. However, patients with cancer usually experience multiple and complex symptoms or disease onsets, and studies or analysis methods addressing multiple concurrent diseases are still very limited. In this study, we proposed a novel algorithm that can analyze multiple outcome variables, and the findings provide clinical implications for clinicians that are not solely based on the results of univariate analyses.



TABLE 2: Present results compared with those of two other studies.

Variable	Chang et al. [12] <sup>a</sup>	Chang et al. [14] <sup>b</sup>	This study
Outcome variables:			
Depression disorders	✓		•
Fracture		✓	•
Osteoporosis			•
Comorbidities:			
Diabetes mellitus (DM)			□
Hypertension			□
Heart failure			□
Coronary heart diseases (CHD)			□
Cerebrovascular diseases			□
Autoimmune diseases			□
Kidney diseases			△
Renal diseases			△
Liver cirrhosis			•
COPD			•
Distant metastatic			•
Primary metastatic			•
Adjuvant therapies:			
Chemotherapy	*	*	•
Radiotherapy	*	*	•
Tamoxifen	*	*	•
AIs	*	*	•
Trastuzumab	*	*	•

<sup>a</sup>[12]. <sup>b</sup>[14]. \* Statistically significant in studies; •, △, and □: implied variables were in the same cluster.

## Abbreviations

MCA: Multiple correspondence analysis  
EM: Expectation-maximization  
SD: Standard deviation  
NHIRD: National Health Insurance Research Database  
NHIA: National Health Insurance Administration  
COPD: Chronic obstructive pulmonary disease  
AI: Aromatase inhibitor  
MOHW: Ministry of Health and Welfare.

## Data Availability

Due to the Personal Information Protection Act of Taiwan, the National Health Insurance Research Database (NHIRD) has been prohibited from releasing insured's medical claim data via applications since June 28, 2016 ([https://nhird.nhri.org.tw/apply\\_00.html](https://nhird.nhri.org.tw/apply_00.html)). The Ministry of Health and Welfare (MOHW) of Taiwan decided that all NHIRD data analyses must be processed in the Health Data Science Center, which was established by the MOHW. Researchers who are interested in NHIRD research can seek access to the NHIRD by formal application (application website: <https://dep.mohw>

.gov.tw/DOS/np-2497-113.html). All the authors of this paper understand and appreciate the need for data transparency in research and are ready to make the data available to those who have a research interest in accessing NHIRD data. Please direct your queries to NHIRD administrators Tze-Hui Wu ([stcarolwu@mohw.gov.tw](mailto:stcarolwu@mohw.gov.tw)) or Zong-Ying Lin ([st-zylin@mohw.gov.tw](mailto:st-zylin@mohw.gov.tw)).

## Conflicts of Interest

The authors declare no conflicts of interest.

## Authors' Contributions

Chieh-Yu Liu was the principal investigator in charge of this study, conceptualizing the study design, applying the study database, coding the algorithm, conducting data analysis, drafting the initial manuscript, and critically reviewing and approving the final, submitted version of the manuscript. Chun-Hung Chang was the research assistant of this study, providing technological support for database maintenance and administrative support. All the authors read and approved the final manuscript.

## Acknowledgments

The authors thank the Health Data Science Center of the Ministry of Health and Welfare of Taiwan for providing access to the National Health Insurance Research Database for data analysis in this study. This manuscript was edited by Wallace Academic Editing.

## References

- [1] C. McKeon, "Reducing the side effects of chemotherapy," *Australian Nursing Journal*, vol. 19, no. 8, p. 41, 2012.
- [2] J. Xie, L. Xu, X. Xu, and Y. Huang, "Complications of peripherally inserted central catheters in advanced cancer patients undergoing combined radiotherapy and chemotherapy," *Journal of Clinical Nursing*, vol. 26, no. 23-24, pp. 4726-4733, 2017.
- [3] R. K. Mittapalli, X. Liu, C. E. Adkins et al., "Paclitaxel-hyaluronic nanoconjugates prolong overall survival in a preclinical brain metastases of breast cancer model," *Molecular Cancer Therapeutics*, vol. 12, no. 11, pp. 2389-2399, 2013.
- [4] S. Tilstra and M. McNeil, "New developments in breast cancer screening and treatment," *Journal of Women's Health*, vol. 26, no. 1, pp. 5-8, 2017.
- [5] T. Li, C. Mello-Thoms, and P. C. Brennan, "Descriptive epidemiology of breast cancer in China: incidence, mortality, survival and prevalence," *Breast Cancer Research and Treatment*, vol. 159, no. 3, pp. 395-406, 2016.
- [6] A. Di Sibio, G. Abriata, D. Forman, and M. S. Sierra, "Female breast cancer in Central and South America," *Cancer Epidemiology*, vol. 44, pp. S110-S120, 2016.
- [7] I. A. Aksenova, M. A. Moore, and A. S. Domozirova, "Trends in breast cancer epidemiology in Chelyabinsk region, 2006-2015," *Asian Pacific Journal of Cancer Prevention*, vol. 18, no. 4, pp. 1163-1168, 2017.
- [8] F. C. Liu, H. T. Lin, C. F. Kuo, L. C. See, M. J. Chiou, and H. P. Yu, "Epidemiology and survival outcome of breast cancer

- in a nationwide study,” *Oncotarget*, vol. 8, no. 10, pp. 16939–16950, 2017.
- [9] N. F. Gomes-Rochette, L. S. Souza, B. O. Tommasi et al., “Association of PvuII and XbaI polymorphisms on estrogen receptor alpha (ESR1) gene to changes into serum lipid profile of post-menopausal women: effects of aging, body mass index and breast cancer incidence,” *PLoS One*, vol. 12, no. 2, article e0169266, 2017.
- [10] M. F. Barginear, H. Muss, G. Kimmick et al., “Breast cancer and aging: results of the U13 conference breast cancer panel,” *Breast Cancer Research and Treatment*, vol. 146, no. 1, pp. 1–6, 2014.
- [11] R. K. C. Yuen, D. Merico, M. Bookman et al., “Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder,” *Nature Neuroscience*, vol. 20, no. 4, pp. 602–611, 2017.
- [12] C. H. Chang, S. J. Chen, and C. Y. Liu, “Adjuvant treatments of breast cancer increase the risk of depressive disorders: a population-based study,” *Journal of Affective Disorders*, vol. 182, pp. 44–49, 2015.
- [13] F. A. Trémollières, I. Ceausu, H. Depypere et al., “Osteoporosis management in patients with breast cancer: EMAS position statement,” *Maturitas*, vol. 95, pp. 65–71, 2017.
- [14] C. H. Chang, S. J. Chen, and C. Y. Liu, “Fracture risk and adjuvant therapies in young breast cancer patients: a population-based study,” *PLoS One*, vol. 10, no. 6, article e0130725, 2015.
- [15] C. H. Chang, S. J. Chen, and C. Y. Liu, “Risk of developing depressive disorders following hepatocellular carcinoma: a nationwide population-based study,” *PLoS One*, vol. 10, no. 8, article e0135417, 2015.
- [16] M. J. Yeh and H. H. Chang, “National Health Insurance in Taiwan,” *Health Affairs*, vol. 34, no. 6, p. 1067, 2015.
- [17] D. Ayele, T. Zewotir, and H. Mwambi, “Multiple correspondence analysis as a tool for analysis of large health surveys in African settings,” *African Health Sciences*, vol. 14, no. 4, pp. 1036–1045, 2014.
- [18] S. N. Kadir, D. F. M. Goodman, and K. D. Harris, “High-dimensional cluster analysis with the masked EM algorithm,” *Neural Computation*, vol. 26, no. 11, pp. 2379–2394, 2014.
- [19] G. J. McLachlan and P. N. Jones, “Fitting mixture models to grouped and truncated data via the EM algorithm,” *Biometrics*, vol. 44, no. 2, pp. 571–578, 1988.
- [20] G. Govaert and M. Nadif, “An EM algorithm for the block mixture model,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 4, pp. 643–647, 2005.
- [21] X. Guo, J. Xu, Y. E, Z. Yu, and T. Sun, “Correlation between hormone receptor status and depressive symptoms in patients with metastatic breast cancer,” *Oncotarget*, vol. 8, no. 31, pp. 50774–50781, 2017.
- [22] A. S. Keuroghlian, L. D. Butler, E. Neri, and D. Spiegel, “Hypnotizability, posttraumatic stress, and depressive symptoms in metastatic breast cancer,” *The International Journal of Clinical and Experimental Hypnosis*, vol. 58, no. 1, pp. 39–52, 2010.
- [23] H. T. Sorensen, S. Friis, J. H. Olsen et al., “Risk of breast cancer in men with liver cirrhosis,” *The American Journal of Gastroenterology*, vol. 93, no. 2, pp. 231–233, 1998.
- [24] J. Collazos, J. Genolla, and A. Ruibal, “Breast cancer-associated antigen CA 15.3 in liver cirrhosis,” *Acta Oncologica*, vol. 31, no. 7, pp. 741–744, 1992.

