Joint Modeling of Estradiol Levels and Survival Data in Breast Cancer Patients: A Case-Cohort Design Study (Postprint)

Authors: Wu Mengjuan, Zhang Tao, Gao Chunjie, Zhao Ting, Wang Lei, Wang Lei

Date: 2024-09-11T00:00:00+00:00

Abstract

Abstract

Background
Breast cancer is a sex hormone receptor-dependent malignant tumor, and dynamic changes in estradiol (E2) play a very important role in the development of breast cancer; classical case-cohort designs completely ignore information from unselected samples, which can easily lead to estimation bias.

Objective
To investigate the impact of dynamic changes in E2 levels on survival prognosis in breast cancer patients and to evaluate the superiority of a modified case-cohort design.

Methods
A total of 8226 patients pathologically diagnosed with breast cancer at the Affiliated Tumor Hospital of Xinjiang Medical University from 2015 to 2019 were selected for follow-up, with the time of diagnosis as the starting point, death due to breast cancer as the endpoint event, and the follow-up deadline being December 31, 2021. Demographic characteristics, immunohistochemical indicators, clinicopathological features, and survival status were collected, and longitudinal measurements of serum E2 levels were performed. Based on the classical case-cohort design, the modified case-cohort design was developed by incorporating survival data from patients outside the case-cohort sample. Under both classical and modified case-cohort designs, linear mixed-effects models and Cox proportional hazards models were used to fit longitudinal data (longitudinal submodel) and survival data (survival submodel) of breast cancer patients, respectively, and a joint model of longitudinal and time-to-event data was established; furthermore, Markov Chain Monte Carlo algorithms were used to

estimate the parameters of the joint model; additionally, the discrimination and calibration of the joint models under classical and modified case-cohort designs were compared using the area under the receiver operating characteristic curve (AUC) and prediction error (PE).

Results
Based on inclusion and exclusion criteria, a total of 895 breast cancer patients were included in the full cohort of this study, of whom 53 died from breast cancer. The median follow-up time was approximately 28 months. One-quarter of the patients were randomly selected from the full cohort as a random subcohort and combined with patients who died during follow-up outside the random subcohort to form the sample for the classical case-cohort design, which included survival data from 236 patients and 1062 measurements of E2 levels. Additionally, based on the classical case-cohort design, survival data from breast cancer patients who remained alive during follow-up outside the classical case-cohort sample (G4) were incorporated as the sample for the modified case-cohort design (including survival data from 895 patients and 1062 E2 level measurements from 236 patients, with 2958 longitudinal missing values of E2 level measurements considered). The joint model results under both classical and modified case-cohort designs showed that dynamic changes in E2 levels were influencing factors for the prognosis of breast cancer patients, and for each unit increase in lg(E2) longitudinally, the risk of death would increase by 23% (HR=1.23, R^=1.015) and 8% (HR=1.08, R^=1.020), respectively. Furthermore, the joint model under the modified case-cohort design demonstrated better discrimination and calibration (AUC=0.706~0.962, PE=0.0012~0.0108).

Conclusion
Longitudinal elevation of E2 levels in breast cancer patients may lead to decreased survival probability. The joint model under case-cohort design can analyze longitudinal and survival data simultaneously, and the modified case-cohort design is superior to the classical case-cohort design.

Full Text

Joint Modeling of Estradiol Levels and Survival Data of Breast Cancer Patients in the Case-Cohort Design

Mengjuan Wu[1], Tao Zhang[1], Chunjie Gao[1], Ting Zhao[2], Lei Wang[3]*

[1]College of Public Health, Xinjiang Medical University, Urumqi 830017, China
[2]Department of Medical Record Management, the Affiliated Cancer Hospital of Xinjiang Medical University, Urumqi 830011, China
[3]Department of Medical Engineering and Technology, Xinjiang Medical University, Urumqi 830017, China

Corresponding author: Lei Wang, Professor; E-mail: wlei81@126.com

Abstract

Background: Breast cancer is a hormone receptor-dependent malignant tumor, and dynamic changes in estradiol (E2) play a critical role in its development. The classical case-cohort design completely ignores information from non-selected samples, which can easily lead to biased estimation. **Objective:** To explore the effect of dynamic changes in E2 levels on survival prognosis in breast cancer patients and evaluate the superiority of an improved case-cohort design. **Methods:** We selected 8,226 patients diagnosed with breast cancer by pathological examination at the Affiliated Cancer Hospital of Xinjiang Medical University from 2015 to 2019. Follow-up started at diagnosis, with death due to breast cancer as the outcome event, and ended on December 31, 2021. Demographic characteristics, immunohistochemical indicators, clinicopathological features, and survival status were collected, and serum E2 levels were longitudinally measured. Based on the classical case-cohort design, the improved design incorporated survival data from patients outside the case-cohort sample. Under both designs, linear mixed effects models and Cox proportional hazards models were used to fit longitudinal data (longitudinal submodel) and survival data (survival submodel), respectively, and joint models for longitudinal and time-to-event data were established. Markov chain Monte Carlo algorithms were used to estimate model parameters. The area under the receiver operating characteristic curves (AUC) and prediction errors (PE) were used to compare discrimination and calibration between the two joint models. **Results:** Based on inclusion and exclusion criteria, 895 breast cancer patients were included in the full cohort, of whom 53 died from breast cancer. The median follow-up time was approximately 28 months. From the full cohort, one-quarter of patients were selected as a random subcohort, which combined with all deaths outside the subcohort formed the classical case-cohort sample (236 patients with survival data and 1,062 E2 measurements). The improved case-cohort design additionally included survival data from patients outside the classical sample who survived during follow-up (G4), comprising survival data from 895 patients and 1,062 E2 measurements from 236 patients (with 2,958 longitudinal measurements assumed missing). Both joint models showed that dynamic changes in E2 levels influenced prognosis: each one-unit longitudinal increase in log(E2) increased mortality risk by 23% (HR=1.23, R=1.015) under the classical design and 8% (HR=1.08, R=1.020) under the improved design. The improved design showed better discrimination and calibration (AUC=0.706–0.962, PE=0.0012–0.0108). **Conclusion:** Longitudinal increases in E2 levels may reduce survival probability in breast cancer patients. Joint models under case-cohort design can simultaneously analyze longitudinal and survival data, with the improved case-cohort design being superior to the classical design.

Keywords: Breast cancer; Estradiol; Case-cohort design; Joint model; Survival data

Introduction

Breast cancer is one of the most common malignant tumors, ranking first in incidence among female cancers, with an estimated 2.3 million new female cases globally in 2020. In China, both incidence and mortality continue to rise, seriously threatening women's health. As a hormone-dependent tumor, estrogen primarily stimulates proliferation of mammary epithelial cells, and binding to estrogen receptors promotes rapid tumor growth and inhibits apoptosis, thereby facilitating breast cancer development. Estrogen mainly includes estradiol (E2) and estrone, with E2 having stronger biological activity and being widely used in clinical research.

Clinically, longitudinal measurements of biomarkers are commonly performed to analyze their dynamic effects on disease progression and estimate temporal trends. Survival data, containing information on survival time, event occurrence, and related factors, are also common in clinical research. To simultaneously analyze longitudinal data, survival data, and their potential associations, joint models have been widely applied. For example, Mchunu et al. used joint models to explore the association between longitudinal CD4 counts and mortality risk in TB/HIV patients, while Wang et al. investigated the relationship between dynamic changes in HDL and LDL cholesterol and metabolic syndrome. When event rates are low, patients from hospital or cancer center follow-ups may not fully represent the general population. The classical case-cohort design, which analyzes all outcome events based on simple random sampling, is suitable for large cohorts with low incidence. However, it completely ignores all information from patients outside the case-cohort sample, creating different event rates between the full and case-cohort samples and leading to selection bias in parameter estimation and survival probability assessment.

Baart et al. evaluated the predictive accuracy of joint models under classical and improved case-cohort designs using simulated data and applied the improved design to explore risk factors for acute coronary syndrome. Building on these discussions, this study uses data from breast cancer patients at the Affiliated Cancer Hospital of Xinjiang Medical University (2015–2019) to improve the case-cohort design by incorporating survival information from patients outside the classical sample. We establish joint models under both designs using linear mixed effects and Cox proportional hazards models, estimate parameters via Markov Chain Monte Carlo (MCMC) algorithms, and compare model discrimination and calibration using AUC and prediction error (PE) to evaluate the improved design's performance and provide evidence for breast cancer prevention and control.

Methods

1.1 General Information We selected 8,226 breast cancer patients diagnosed by pathological examination at the Affiliated Cancer Hospital of Xinjiang

Medical University from 2015 to 2019. Follow-up started at diagnosis, with death due to breast cancer as the outcome event, and ended on December 31, 2021. Data were collected through the hospital's electronic medical record and follow-up systems, including demographic characteristics, immunohistochemical indicators, clinicopathological features, survival status, and longitudinal serum E2 measurements (pmol/L). This study was approved by the Medical Ethics Committee of the Affiliated Cancer Hospital of Xinjiang Medical University (Approval No.: K-2023001).

Inclusion criteria: (1) Age >18 years; (2) Primary tumor pathologically diagnosed as breast cancer; (3) Signed informed consent; (4) $ $2 repeated measurements during follow-up. **Exclusion criteria:** (1) Other malignant tumors; (2) <2 repeated measurements; (3) Missing important pathological information. Based on these criteria, 895 patients were included in the final cohort.

1.2.1 Case-Cohort Design Consider a study with n independent individuals. Let T_i^* and C_i^* denote the true event time and censoring time for individual i, respectively, with observed time $T_i = \min(T_i^*, C_i^*)$. The event indicator is $\delta_i = I(T_i^* \leq C_i^*)$, where $I(\cdot)$ is the indicator function. Individuals experiencing the outcome are cases (G2 G3 in [Figure 1: see original paper]), while those not experiencing it are non-cases (G1 G4).

The classical case-cohort design selects a random subcohort via simple random sampling from the full cohort (G1 G2 in [Figure 1: see original paper]). The case-cohort sample comprises this subcohort plus all cases outside it (G3, where $\delta_i = 1$). Indicator variables S_i and CC_i denote selection into the random subcohort and case-cohort sample, respectively:

$$CC_i = \begin{cases} 1, & \text{if } \delta_i = 1 \text{ or } S_i = 1 \\ 0, & \text{if } \delta_i = 0 \text{ and } S_i = 0 \end{cases}$$

This study improves upon the classical design by additionally incorporating survival information from individuals outside the case-cohort sample who survived during follow-up (G4), without considering their longitudinal data—termed the improved case-cohort design.

Let $m_i(t)$ denote the true longitudinal value at time t, with observed value $y_i(t)$. For individuals outside the case-cohort sample (G4), longitudinal data are assumed missing, denoted as $y_i^m(t)$. The observed sets are:

$$\begin{aligned} G1 &= \{y_i, T_i, \delta_i = 0, S_i = 1, CC_i = 1; i = 1, \ldots, n\} \\ G2 &= \{y_i, T_i, \delta_i = 1, S_i = 1, CC_i = 1; i = 1, \ldots, n\} \\ G3 &= \{y_i, T_i, \delta_i = 1, S_i = 0, CC_i = 1; i = 1, \ldots, n\} \\ G4 &= \{y_i^m, T_i, \delta_i = 0, S_i = 0, CC_i = 0; i = 1, \ldots, n\} \end{aligned}$$

1.2.2 Joint Model for Longitudinal and Survival Data The longitudinal submodel uses a linear mixed effects model:

$$y_i(t) = m_i(t) + \varepsilon_i(t) = x_i(t)^T\beta + z_i(t)^T b_i + \varepsilon_i(t)$$

where $y_i(t)$ is the observed longitudinal value, $m_i(t)$ is the true value, $x_i(t)$ and $z_i(t)$ are design matrices for fixed and random effects, β is the fixed effects vector, $b_i \sim N(0, D)$ is the random effects vector, and $\varepsilon_i(t) \sim N(0, \sigma^2)$ is measurement error. We used a linear mixed effects model with natural cubic spline functions of time to capture non-linear temporal trends.

The association between longitudinal dynamics and survival risk is quantified by coefficient α. The survival submodel uses an extended Cox proportional hazards model:

$$h_i(t \mid M_i(t), \omega_i) = h_0(t)\exp\{\omega_i^T\gamma + \alpha m_i(t)\}$$

where $M_i(t) = \{m_i(s), 0 < s < t\}$, $h_0(t)$ is the baseline hazard, ω_i is the design matrix for time-independent covariates, γ are unknown parameters, and α is the association parameter.

Assuming shared random effects b_i between submodels, the joint model is:

$$\begin{aligned} y_i(t) &= m_i(t) + \varepsilon_i(t) = x_i(t)^T\beta + z_i(t)^T b_i + \varepsilon_i(t) \\ h_i(t \mid M_i(t), \omega_i) &= h_0(t)\exp\{\omega_i^T\gamma + \alpha m_i(t)\} \\ b_i &\sim N(0, D), \quad \varepsilon_i(t) \sim N(0, \sigma^2) \end{aligned}$$

We estimated parameters using MCMC algorithms under a Bayesian framework. For individual i, the contribution to the survival likelihood is:

$$p(T_i, \delta_i \mid b_i, \beta, \theta_t) = h_i\{T_i \mid M_i(T_i), \theta_t\}^{\delta_i} S_i\{T_i \mid M_i(T_i), \theta_t\} = [h_0(T_i \mid \gamma_s)\exp\{\gamma^T\omega_i + \alpha m_i(T_i)\}]^{\delta_i} \times \exp\{-\int_0^{T_i} h_($$

where $\theta_t = (\gamma_s, \gamma, \alpha)$ and γ_s are parameters for the baseline hazard B-spline.

For the classical design, individuals in $G1 \cup G2 \cup G3$ have complete longitudinal information, with contribution to the longitudinal likelihood:

$$p(y_i \mid b_i, \theta_y) = \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y_{ij} - x_{ij}^T\beta - z_{ij}^T b_i)^2}{2\sigma^2}\right\}$$

where $\theta_y = (\beta, \sigma)$. For the improved design, individuals in G4 have missing longitudinal data, and their longitudinal submodel posterior is:

$$p(y_i^m \mid T_i, \delta_i = 0, G4) = \int p(y_i^m \mid T_i, \delta_i = 0, \theta_y) p(\theta_y \mid G4) d\theta_y$$

The full joint posterior distributions are:

$$\begin{aligned} \text{Classical: } & p(\theta, b_i \mid T_i, \delta_i, y_i) \propto p(T_i, \delta_i \mid b_i, \theta) p(y_i \mid b_i, \theta) p(b_i \mid \theta) p(\theta) \\ \text{Improved: } & p(\theta, b_i, y_i^m \mid T_i, \delta_i) \propto p(T_i, \delta_i \mid b_i, \theta) p(y_i^m \mid b_i, \theta) p(b_i \mid \theta) p(\theta) \end{aligned}$$

where θ represents all unknown parameters.

We evaluated model discrimination and calibration using AUC and PE. Higher AUC (closer to 1) indicates better discrimination, while smaller PE indicates better calibration (smaller difference between observed and predicted survival rates).

Results

2.1 Data Processing The full cohort included 895 breast cancer patients, of whom 53 died from breast cancer. The median follow-up was approximately 28 months. During follow-up, 4,020 E2 measurements were collected (mean 4.49 per patient). From the full cohort, one-quarter were randomly selected as a subcohort, which combined with deaths outside the subcohort formed the classical case-cohort sample (236 patients with survival data and 1,062 E2 measurements). The improved case-cohort design additionally included survival data from patients outside the classical sample who survived during follow-up (G4), comprising survival data from 895 patients and 1,062 E2 measurements from 236 patients (with 2,958 longitudinal measurements assumed missing). Baseline characteristics are shown in . Normality tests showed E2 levels were skewed; we applied log10 transformation to achieve approximate normality for log(E2).

2.2 Longitudinal and Survival Submodels We fitted log(E2) using a linear mixed effects model with natural cubic spline functions of time. The final longitudinal submodel was:

$$\log(\text{E2}) = y_i(t) = m_i(t) + \varepsilon_i(t) = \beta_1 + \beta_2 t + \beta_3 t^2 + \beta_4 t^3 + b_1 + b_2 t + b_3 t^2 + b_4 t^3 + \varepsilon_i(t)$$

We plotted trajectories for six randomly selected patients ([Figure 2: see original paper]), revealing non-linear patterns (blue dashed lines).

Univariate and multivariate Cox regression identified TNM stage II (HR=3.59, 95%CI=1.05–12.20), III (HR=14.00, 95%CI=4.18–46.70), IV (HR=29.90, 95%CI=7.58–118.00) and HER-2 positivity (HR=2.11, 95%CI=1.19–3.76) as

risk factors, while breast-conserving surgery (HR=0.12, 95%CI=0.01–0.95) was protective (). The optimal survival submodel was:

$$h_i(t) = h_0(t)\exp\{\gamma_1\omega_{\text{TNM-II}}+\gamma_2\omega_{\text{TNM-III}}+\gamma_3\omega_{\text{TNM-IV}}+\gamma_4\omega_{\text{surgery-conserving}}+\gamma_5\omega_{\text{surgery-radical}}+\gamma_6\omega_{\text{HER-2+}}+\alpha m_i($$

The final joint model was:

$$\begin{aligned}\log(\text{E2}) &= \beta_1 + \beta_2 t + \beta_3 t^2 + \beta_4 t^3 + b_1 + b_2 t + b_3 t^2 + b_4 t^3 + \varepsilon_i(t)\\ h_i(t) &= h_0(t)\exp\{\gamma_1\omega_{\text{TNM-II}} + \gamma_2\omega_{\text{TNM-III}} + \gamma_3\omega_{\text{TNM-IV}} + \gamma_4\omega_{\text{surgery-conserving}} + \gamma_5\omega_{\text{surgery-radical}} + \gamma_6\omega_{\text{HER-2-}}\\ b_i &\sim N(0, D), \quad \varepsilon_i(t) \sim N(0, \sigma^2)\end{aligned}$$

2.3 Joint Model Results Parameter estimates are shown in . The association coefficient α converged in both models ($\hat{R}$<1.05). Holding other baseline variables constant, each one-unit longitudinal increase in log(E2) increased mortality risk by 23% (HR=exp(0.21)=1.23, $\hat{R}$=1.015) under the classical design and 8% (HR=exp(0.08)=1.08, $\hat{R}$=1.020) under the improved design.

Predictive accuracy comparisons across time points and intervals ($t+\Delta t$) showed the improved design had higher AUC (0.706–0.962) and lower PE (0.0012–0.0108) than the classical design (AUC=0.693–0.930, PE=0.0013–0.0120) (). For example, at $\Delta t = 5$ and $t = 9$, the classical design yielded AUC=0.797 and PE=0.0059, while the improved design yielded AUC=0.823 and PE=0.0055.

Discussion

The classical case-cohort design reduces bias from lack of repeatability, while the improved design maintains sample size efficiency and avoids selection bias by incorporating all patients' survival information. Joint models effectively monitor longitudinal trajectories. Previous studies have shown significant associations between E2 dynamics and breast cancer prognosis. This study used shared random effects to jointly model linear mixed effects and Cox models under both designs to explore the relationship between E2 dynamics and survival in Xinjiang breast cancer patients.

Cox regression showed advanced TNM stage was an independent prognostic risk factor, consistent with known biology where larger tumors and higher metastasis probability increase mortality risk. Breast-conserving surgery (HR=0.12, 95%CI=0.01–0.95) significantly improved prognosis, aligning with studies by Lagendijk et al. and Veronesi et al. demonstrating that surgery reduces tumor volume and controls dissemination. Both joint models identified elevated E2 as a risk factor, consistent with literature. Kensler et al. found elevated prediagnostic E2 was a mortality risk factor in the Nurses' Health Study. Our

results showed each one-unit increase in log(E2) increased mortality risk by 23% (classical) and 8% (improved). Elevated E2 may increase mammary cell proliferation and reduce endocrine therapy efficacy. Clinical monitoring should prompt comprehensive assessment and treatment adjustment.

The improved design demonstrated better dynamic prediction accuracy across time points and intervals, with higher AUC (0.706–0.962) and lower PE (0.0012–0.0108). This indicates that incorporating the full cohort's survival information improves joint model estimation, reduces selection bias, and enhances cost-effectiveness.

Limitations: The 95%CI for the association coefficient was wide, possibly due to small sample size and limited statistical power. Future studies should expand sample sizes. Additionally, interactions between E2, progesterone, and other hormones were not considered. Future work could incorporate progesterone as a time-varying covariate in a multivariate joint model to optimize predictive accuracy.

Conclusion

Elevated E2 levels are a risk factor for mortality in breast cancer patients. Joint models under case-cohort design can simultaneously analyze longitudinal and survival data, with the improved case-cohort design demonstrating superior predictive performance.

Author Contributions: Mengjuan Wu: conceptualization, data analysis, model construction, coding, visualization, original draft. Tao Zhang: data analysis, visualization. Chunjie Gao: coding, revision. Ting Zhao: data collection, funding. Lei Wang: supervision, conceptual guidance, revision, funding. All authors approved the final manuscript.

Conflicts of Interest: None declared.

ORCID IDs:
Mengjuan Wu: https://orcid.org/0009-0006-4546-410X
Lei Wang: https://orcid.org/0000-0003-4635-8879

References

[1] SUNG H, FERLAY J, SIEGEL R L, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA Cancer J Clin, 2021, 71(3): 209-249. DOI: 10.3322/caac.21660.

[2] LEI S Y, ZHENG R S, ZHANG S W, et al. Breast cancer incidence and mortality in women in China: temporal trends and projections to 2030[J]. Cancer

Biol Med, 2021, 18(3): 900-909. DOI: 10.20892/j.issn.2095-3941.2020.0523.

[3] Chen Xianrong, Xiao Bingrong, Cheng Hao. Clinical Analysis of Combined Detection of B-ultrasound, Thyroid Hormones, and Estradiol Levels in Breast Cancer Patients[J]. Chinese Journal of Health Laboratory Technology, 2019, 29(10): 1245-1247.

[4] Wang Juanjuan, Gu Yunong, Li Jian, et al. Study on the Correlation Between Sex Hormone Levels and Female Breast Cancer[J]. Journal of Clinical and Pathological Research, 2015, 35(6): 1096-1102. DOI: 10.3978/j.issn.2095-6959.2015.06.044.

[5] ZHU D T, ZHAO Z, CUI G M, et al. Single-cell transcriptome analysis reveals estrogen signaling coordinately augments one-carbon, polyamine, and purine synthesis in breast cancer[J]. Cell Rep, 2019, 27(9): 2798. DOI: 10.1016/j.celrep.2019.05.047.

[6] Mao Xu. Research Progress on the Effect of CYP1-Mediated Estradiol Bioactivation on the Development of Female Breast Cancer[J]. Chinese Journal of Pharmacology and Toxicology, 2023, 37(8): 640-644. DOI: 10.3867/j.issn.1000-3002.2023.08.009.

[7] Zhai Yinghong, Chen Qi, Han Hedong, et al. Introduction to Joint Models and Their Application in Medical Research[J]. Chinese Journal of Epidemiology, 2019, 40(11): 1456-1460. DOI: 10.3760/cma.j.issn.0254-6450.2019.11.021.

[8] MCHUNU N N, MWAMBI H G, RIZOPOULOS D, et al. Using joint models to study the association between CD4 count and the risk of death in TB/HIV data[J]. BMC Med Res Methodol, 2022, 22(1): 295. DOI: 10.1186/s12874-022-01775-7.

[9] Xie Weihua, Yu Xiaojin, Dai Pinyuan, et al. Application of Bayesian Joint Models in Studying the Association Between Pulse Pressure Changes and All-Cause Mortality in the Elderly[J]. Chinese Journal of Disease Control & Prevention, 2021, 25(1): 72-77. DOI: 10.16462/j.cnki.zhjbkz.2021.01.014.

[10] Wang Xiaorong, Liu Ying, Li Meng, et al. Study on the Correlation Between Dynamic Changes in Lipoprotein Cholesterol and Metabolic Syndrome[J]. China Preventive Medicine Journal, 2018, 19(6): 443-447. DOI: 10.16506/j.1009-6639.2018.06.009.

[11] BAART S J, BOERSMA E, RIZOPOULOS D. Joint models for longitudinal and time-to-event data in a case-cohort design[J]. Stat Med, 2019, 38(12): 2269-2281. DOI: 10.1002/sim.8113.

[12] PRENTICE R L. A case-cohort design for epidemiologic cohort studies and disease prevention trials[J]. Biometrika, 1986, 73(1): 1-11. DOI: 10.1093/biomet/73.1.1.

[13] FENG Y, BAI Y S, LU Y J, et al. Plasma perfluoroalkyl substance exposure and incidence risk of breast cancer: a case-cohort study in the

Dongfeng-Tongji cohort[J]. Environ Pollut, 2022, 306: 119345. DOI: 10.1016/j.envpol.2022.119345.

[14] YAO S, KWAN M L, ERGAS I J, et al. Association of serum level of vitamin D at diagnosis with breast cancer survival: a case-cohort analysis in the pathways study[J]. JAMA Oncol, 2017, 3(3): 351-357. DOI: 10.1001/jamaoncol.2016.4188.

[15] ANDRINOPOULOU E R, HARHAY M O, RATCLIFFE S J, et al. Reflection on modern methods: dynamic prediction using joint models of longitudinal and time-to-event data[J]. Int J Epidemiol, 2021, 50(5): 1731-1743. DOI: 10.1093/ije/dyab047.

[16] LUO Y, LI H B, ZHANG Y, et al. Combination of endogenous estradiol and adipokine leptin in breast cancer risk and prognosis assessment in postmenopausal Chinese women[J]. Front Endocrinol, 2021, 12: 766463. DOI: 10.3389/fendo.2021.766463.

[17] SAVAGE P, YU N, DUMITRA S, et al. The effect of the American Joint Committee on Cancer eighth edition on breast cancer staging and prognostication[J]. Eur J Surg Oncol, 2019, 45(10): 1817-1820. DOI: 10.1016/j.ejso.2019.03.027.

[18] LAGENDIJK M, VAN MAAREN M C, SAADATMAND S, et al. Breast conserving therapy and mastectomy revisited: breast cancer-specific survival and the influence of prognostic factors in 129,692 patients[J]. Int J Cancer, 2018, 142(1): 165-175. DOI: 10.1002/ijc.31034.

[19] VERONESI U, CASCINELLI N, MARIANI L, et al. Twenty-year follow-up of a randomized study comparing breast-conserving surgery with radical mastectomy for early breast cancer[J]. N Engl J Med, 2002, 347(16): 1227-1232. DOI: 10.1056/NEJMoa020989.

[20] KENSLER K H, ELIASSEN A H, ROSNER B A, et al. Pre-diagnostic sex hormone levels and survival among breast cancer patients[J]. Breast Cancer Res Treat, 2019, 174(3): 749-758. DOI: 10.1007/s10549-018-05121-8.

[21] ARTHUR R S, XUE X N, ROHAN T E. Prediagnostic circulating levels of sex steroid hormones and SHBG in relation to risk of ductal carcinoma in situ of the breast among UK women[J]. Cancer Epidemiol Biomarkers Prev, 2020, 29(5): 1058-1066. DOI: 10.1158/1055-9965.EPI-19-1302.

[22] LI J X, LI C Y, FENG Z W, et al. Effect of estradiol as a continuous variable on breast cancer survival by menopausal status: a cohort study in China[J]. Breast Cancer Res Treat, 2022, 194(1): 103-111. DOI: 10.1007/s10549-022-06593-5.

[23] EL-ATTAR A A, IBRAHIM O M, ALHASSANIN S A, et al. Effect of metformin as an adjuvant therapy to letrozole on estradiol and other biomarkers involved in the pathogenesis of breast cancer in overweight and obese post-

menopausal women: a pilot study[J]. Eur J Clin Pharmacol, 2023, 79(2): 299-309. DOI: 10.1007/s00228-022-03444-6.

[24] Dang Hong. Research on Confidence Intervals and Credible Intervals[J]. Studies in College Mathematics, 2023, 26(1): 47-50. DOI: 10.3969/j.issn.1008-1399.2023.01.017.

[25] Huang Shen, Jiang Qingqing, Wang Shiqi, et al. P-values and Confidence Intervals: Connections and Differences, Misuse and Controversies[J]. Journal of Mathematical Medicine, 2023, 36(1): 3-8. DOI: 10.12173/j.issn.1004-4337.202212021.

[26] VAN OUDENHOVEN F M, SWINKELS S H N, HARTMANN T, et al. Modeling the underlying biological processes in Alzheimer's disease using a multivariate competing risk joint model[J]. Stat Med, 2022, 41(17): 3421-3433. DOI: 10.1002/sim.9425.

[27] TANG A M, ZHAO X Q, TANG N S. Bayesian variable selection and estimation in semiparametric joint models of multivariate longitudinal and survival data[J]. Biom J, 2017, 59(1): 57-78. DOI: 10.1002/bimj.201500070.

Received: December 10, 2023; Revised: July 13, 2024
Edited by: Jia Mengmeng

Note: Figure translations are in progress. See original paper for figures.