Identification of Breast-Cancer-Associated Properties of Drinking Water under a Composite-Toxicity Perspective of Mixed Contaminants: A Case Study in a High-Prevalence Area of China

Shuhan Peng 1,2,3, Shengkun Dong 1,2,3, Chang Gong 4, Xiaohong Chen 1,2,3,*, Hongyu Du 1, Yuehao Zhan 1 and Changxin Ye 1,2,3

1 School of Civil Engineering, Sun Yat-sen University, Guangzhou 510275, China
2 Guangdong Engineering Technology Research Center of Water Security Regulation and Control for Southern China, Sun Yat-sen University, Guangzhou 510275, China
3 Key Laboratory of Water Cycle and Water Security in Southern China of Guangdong High Education Institute, Sun Yat-sen University, Guangzhou 510275, China
4 Breast Tumor Center, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou 510120, China

* Correspondence: eescxh@mail.sysu.edu.cn

Abstract: Breast cancer is the most frequently diagnosed female cancer worldwide. Environmental contaminant exposure is suspected to be crucial, but the broad-spectrum communal properties that these suspected contaminants all share remain to be explored, especially in source and drinking water. In this work, we focused on the Pearl River Basin, which has the highest breast cancer incidence and mortality in China, and hypothesized that the breast cancer risk in this area is associated with its water source. Our objective was to resolve the possible communal properties that are associated with breast cancer from water mixture extracts of source and drinking water and to identify the key drivers by utilizing the latest epidemiology data, performing an exhaustive water toxicological and chemical characterization, and combining partial least-squares path statistics modeling (PLS-PM). We proposed a path for a drinking water-toxicity-induced breast cancer risk and confirmed its association with estrogen-receptor- and thiol-depletion-relevant mechanisms. The breast cancer incidence risk was associated with water-mixture-promoted mammalian cell proliferation (i.e., estrogenic effect), while the mortality risk was associated with a greater thiol depletion (i.e., oxidative stress). Endocrine-disrupting chemicals (EDCs) and dissolved organic matter (DOM) from anthropogenic sources in drinking water are key drivers for estrogenic effects and oxidative stress, respectively. The PLS-PM standardized effects of the DOM and EDCs in treated water on the breast cancer incidence and mortality were −0.07 and 0.31, and 0.35 and 0.31, respectively, further revealing that EDCs strongly influence the incidence risk, whereas the mortality risk resulted from the joint effects of EDCs and DOM. This study clearly shows an association between the breast cancer risk and drinking water toxicity in a high-prevalence area of China, broadening the future perspectives for water-contaminant-specific breast cancer prevention research.

Keywords: breast cancer; drinking water; estrogenic effect; oxidative stress; endocrine-disrupting chemicals; dissolved organic matter

1. Introduction

Among the female population, breast cancer is currently the most frequently diagnosed cancer and the leading cause of cancer death, ranking first in incidence and mortality in most countries [1]. Unfortunately, although environmental factors are suspected to play a more important role than genetic factors, about 50% of female breast cancer cases do not have specific known risk factors [2,3]. The geographic heterogeneity exhibited by global breast cancer clusters has raised particular attention to the link between environmental contaminants and the breast cancer risk [4,5]. As a critical medium connecting human
and environmental exposure, the quality of drinking water is associated with the total cancer incidence and the mortality risk [6,7]. However, the ongoing uncertainty about the relationship between water contaminants and breast cancer has hampered global prevention efforts.

Since the discovery of spatial associations between contaminated drinking water and breast cancer clusters [8], many studies have attempted to identify the potential key water contaminants. However, due to a lack of patient data and water contaminant measurements, these early studies produced only weak correlations with limited water contaminants, such as plumbing leachates [9] and nitrates [10]. With the increasing diversity of water contaminants, it has also become more complicated to explore the link to the breast cancer risk at the level of individual contaminants. Together, these limitations have prevented a comprehensive delineation of the risk factors inherent in drinking water and impeded the resolution of the broad-spectrum communal breast-cancer-associated properties shared by these suspected contaminants. Recently, several breast cancer case-control studies have shown elevated levels of certain endocrine-disrupting chemicals (EDCs) in the breast cancer population, such as nonylphenols [11], phthalates [12], and certain metals [13]. However, it remains uncertain whether their sources include drinking water. Meanwhile, ubiquitous dissolved organic matter (DOM) represents another primary risk source in drinking water [14]. DOM is a precursor to disinfection by-products (DBPs), which have been associated with specific cancer risks (i.e., bladder cancer, colon cancer, and rectal cancer) [15]. More importantly, DOM can alter the toxicological properties of water by interacting with multiple pollutants (including EDCs [16,17]), exhibiting genotoxicity and estrogenic/anti-estrogenic activities [18,19]. Therefore, it is insufficient to identify the breast-cancer-associated properties inherent in drinking water solely from the perspective of EDCs without considering DOM. Unfortunately, due to DOM’s high complexity and diversity, its impact on the breast cancer risk remains unclear.

Toxicological studies based on single-contaminant exposures have shown that specific EDCs, such as nonylphenol, can disrupt the mammalian endocrine system, causing abnormal cell proliferation by interfering with estrogen receptor (ER)-mediated signaling and increasing the breast cancer risk [20,21]. On the other hand, some DOM components alter the cellular oxidative stress levels and lead to unintended cell apoptosis by inducing the production of reactive oxygen species [22]. Notably, imbalances in estrogen signaling and oxidative stress are commonly observed in breast cancer [20,23]. However, due to the coexistence and interaction of multiple contaminants in the environment, synergistic and antagonistic toxicity effects are inevitable; the toxicological effects of mixed contaminants may differ significantly from those of individual contaminants. Therefore, toxicity assessments based on mixed contaminants allow for a better approximation of the actual exposure.

In general, the current research on the environmental etiology of breast cancer by targeting individual-contaminant exposure remains limited in terms of the examined contaminant species and faces challenges in reflecting the toxicological effects under combined-exposure scenarios. Moreover, given the wide variety of environmental contaminants, extracting the shared characteristics among these contaminants associated with breast cancer would facilitate a more comprehensive understanding of the unclear environmental etiology of breast cancer.

CHO cells from Chinese hamster ovaries are known for their high expression of multiple ERs, and estrogen-like components modulate their cell cycle by inducing the expression of ER-related genes [24]. The cysteine thiol is the primary cellular reductant defense against oxidative stress [25]. The depletion of the thiol pool may result in a wide range of adverse biological responses associated with oxidative stress, which is directly linked to breast cancer [25–27]. Therefore, CHO cells and biological thiols, capable of reflecting mixed estrogenic activity and the overall oxidative stress levels, are ideal for characterizing the composite toxicity effects related to breast cancer following a drinking-water-mixture exposure.
The objective of this study was to resolve possible communal properties associated with breast cancer from the water mixture extracts of varying contaminant compositions and to identify the key drivers. Field sampling was conducted in the Pearl River Basin (PRB), which covers a population of nearly 5 million people and is a region in China with one of the highest prevalences of breast cancer [28,29]. To quantify the ER interference and oxidative stress induction of the water mixture contaminants, as reflected by mammalian cell growth stimulation and thiol depletion, extensively used CHO cell cytotoxicity and thiol-specific reactivity analyses were performed as two toxicity endpoints. A partial least-squares path modeling (PLS-PM) analysis, which is widely applied to examine the complex interactions between human and environmental factors [15], was further performed to investigate the proposed breast-cancer-risk-inducing path hypothesis based on the resolved toxicological commonalities of drinking water mixture contaminants. The relative impacts of DOM and EDCs on the breast cancer risk were quantified based on path coefficients. Exploring the underlying mechanism of drinking water contaminants on the breast cancer risk could broaden our understanding of its unclear etiology and help develop effective prevention strategies.

2. Materials and Methods

2.1. Breast Cancer Prevalence in the Study Area

The PRB is the second largest runoff basin among China’s nine major basins. According to statistics from the National Cancer Center, the PRB has the highest breast cancer incidence and mortality in China, which are over 20–25% higher than China’s average, reaching 34.8/10^5 and 7.6/10^5 (age-standardized by Segi’s population), respectively [30]. Intriguingly, breast cancer patterns in the PRB exhibit geographic heterogeneity [29]: certain regions have a higher incidence than China’s or even the world’s averages (46.3–47.8/10^5), while neighboring regions have a lower incidence. Meanwhile, the breast cancer clusters in this area are geographically associated with uterus and ovarian cancers, which is consistent with other breast-cancer-prevalent areas worldwide and collectively points to environmental risk factors [3]. Independent sources of drinking water have long been suspected as a potential factor in this pattern. Therefore, the densely populated lower reaches of the PRB were selected to explore the relationship between drinking water and breast cancer (Figure S1). The populations in these 11 counties are mostly natives (i.e., born and lived locally) who share similar dietary habits and use the same Cantonese dialect/language, with negative migration rates over a decade.

To better assess the average level of breast cancer in each study zone, we covered the most extended possible period of cancer data, since 2014, the earliest year for which cancer data were available. The most complete data on the breast cancer incidence rates (age-standardized by Segi’s population) available for the study area were from 2014 to 2016, and were obtained from the China Cancer Registry Annual Report [31]. The most complete data on the mortality rates (age-standardized by Chinese population) were from 2014 to 2020, and were obtained from the Guangdong Provincial Center for Disease Control and Prevention via a data request (cdcp.gd.gov.cn). It should be noted that these epidemiologic cancer data are population-based, do not involve any individual patients, and were assessed for data quality by the National Cancer Center before publication. Each zone’s multi-year average age-standardized incidence and mortality rates were regarded as the risk of breast cancer incidence and mortality (Figure 1). Since the registration of cancer incidence is not yet available in most zones, four zones provided both incidence and mortality data (i.e., zone 2/7/10/11, Figure 1a), while the remaining zones provided only mortality data.
2.2. Water Sample Collection

Before the collection of water samples, we carefully read the sampling precautions declared in the test standards (i.e., China standard) for the corresponding contaminants to ensure that the sampling process complied with the specifications. Also, standard chemicals were used to validate the determination methods in the test standards, and the corresponding standard curves were established in advance to ensure the accuracy of contaminant identification and quantification.

In brief, four rounds of sampling were conducted at 17 drinking water treatment plants (DWTPs) in the 11 zones (March and November 2021 and May 2022). The overview of each DWTP and the sampling locations are detailed in Table S1 and Figure S2. These DWTPs all adopted the conventional process of coagulation–sedimentation–filtration–chlorine.
disinfection. The water source had not experienced precipitation for 48 h before the water sampling. For each DWTP, 15 L of raw water and 15 L of treated water were collected per round using pre-washed dark-colored glass bottles. All the samples were filtered through 0.45 µm glass fiber filters and stored at 2 °C for further analyses.

2.3. Toxicological Analysis of Water Samples

The mammalian cell cytotoxicity and thiol-specific reactivity were used as two typical toxicity endpoints to assess the estrogenic activity and oxidative stress levels induced by the water mixtures. These two toxicological test methods are detailed in Supplementary Text S1 and Supplementary Text S2, respectively. Before the toxicity analysis, the water mixture contaminants were first extracted from 5 L of the collected samples using a solid-phase extraction column (Oasis HTE, Waters, Milford, MA, USA). The column was then eluted with 10 mL of methanol and 5 mL of dichloromethane. The eluent was concentrated to a final volume of 0.5 mL in dimethyl sulfoxide at a concentration factor of 1 × 10⁵ to produce the concentrated water sample (CWS), which was sealed at −20 °C for further use.

The mammalian cell cytotoxicity analysis was performed in the Chinese hamster ovary (CHO) cell line K1.AS52 (clone 11–4–8) [32]. For each experimental group, CHO cells were exposed to a range of CWSs in sealed 96-well microplates during a 72 h incubation. Following histological staining with crystal violet, the cell viability in each microwell was quantified at an absorbency of 595 nm. The absorbance of each CWS was normalized to the negative control and concentration–response curves were generated. A regression analysis was applied to determine the LC₅₀ value for each curve, and the CHO cytotoxicity index (CTI) was the LC₅₀⁻¹ × 10³.

The thiol-specific reactivity analysis assessed the depletion of thiols, which are supplied by N-acetylcysteine (NAC), after exposure to the CWSs [32]. For each experimental group, a range of CWSs in sealed 96-well microplates was reacted with NAC for 20 min. Concurrently, the microwells containing series dilutions of CWSs without adding NAC were set as sample blanks. Microwells with maleimide were set as positive controls, and microwells without any toxins were set as negative controls. Following the addition of 5,5-dithiobis (2-nitrobenzoic acid), the remaining thiols in each microwell were quantified at an absorbency of 412 nm. After a blank correction, the percentage of microplate A₄₁₂ values for each experimental group and the positive control relative to the concurrent negative control were determined. Concentration–response curves were generated for each CWS using these normalized data. A regression analysis was applied to determine the EC₅₀ value for each curve, and the thiol-specific reactivity index (TRI) value was the EC₅₀⁻¹ × 10³.

2.4. Characterization of DOM

The total organic carbon (TOC) of the water samples was determined using a TOC-LCPH analyzer (Shimadzu, Kyoto, Japan). Ultraviolet–visible spectroscopy (UV-vis) was applied to characterize the molecular structure of DOM. The samples were measured using a UV2300II spectrophotometer (Techcomp, Hong Kong, China) and blank-corrected using ultrapure water. Several widely used UV-vis parameters were calculated for general DOM characterization, as detailed in Figure S3.

Three-dimensional excitation–emission matrix fluorescence spectroscopy (EEM) was used to characterize the species and content of DOM with fluorescent properties. The samples were measured using an F-7000 fluorescence spectrophotometer (Hitachi, Tokyo, Japan). The spectra’s Rayleigh and Raman scattering regions were eliminated and interpolated by the Delaunay triangulation method [33]. Before each measurement batch, ultrapure water was scanned for the blank correction and Raman normalization. Component separation of the mixed fluorescence signals of DOM was performed by the widely-used parallel factor (PARAFAC) modeling and identified by the OpenFlour database (openfluor.com) [34].

Fourier-transform infrared spectroscopy (FTIR) was applied to characterize the functional groups of DOM. An amount of 1.0 mg of freeze-dried sample powder was mixed and
ground with 100 mg of dried potassium bromide. The FTIR spectra were collected using a Nicolet iS50 FT-IR spectrometer (Thermo Fisher, Waltham, MA, USA). After a background and baseline correction, the FTIR spectra of each sample were obtained.

2.5. Measurement of EDCs

Nonylphenol [20], arsenic (As) [35], antimony (Sb), barium (Ba) [36], cobalt (Co), nickel (Ni) [37], boron (B) [38], nitrate [39], and fluoride [40] have been reported to influence the mammalian endocrine system, and they may be related to breast cancer. The determination of nonylphenol was performed by solid-phase extraction and high-performance liquid chromatography (China standard HJ 1192-2021), which are detailed in Supplementary Text S3. The soluble elements As, Sb, Ba, Co, Ni, and B were determined by inductively coupled plasma–mass spectrometry (China standard HJ 700-2014). Nitrate and fluoride were determined by ion chromatography (China standard HJ 84-2016).

2.6. Screening Process of Key Water Contaminants

As the World Health Organization suggests, a correlation analysis is a practical statistical method for finding environmental risk factors for breast cancer [3]. To identify the key contaminants among water mixtures and to explore their broad-spectrum communal properties associated with breast cancer, the following screening process was applied using this statistical method. Pearson correlation analyses were first performed between the examined DOM and EDC species with the breast cancer incidence and mortality. Contaminants with no noticeable correlation between either the incidence or mortality were excluded (Pearson’s r < 0.4). The remaining contaminants were then correlated with water toxicity indexes (CTI and TRI) to further screen for potential drivers of water mixture toxicity. Finally, toxicological or epidemiological evidence from other studies for the screened contaminants was explored to minimize potential pseudo-correlation results. All the tests were two-sided with \( p < 0.05 \) being considered statistically significant. For the zone with several DWTPs, the contaminant contents and toxicity indexes were averaged over the samples from multiple DWTPs.

2.7. PLS-PM Analysis

A PLS-PM analysis has been proven to be useful in quantifying the toxicological associations of drinking water components in other cancer environmental etiology studies [15]. Therefore, this statistical method was also employed in this study to analyze the association of water components with water toxicity and breast cancer. The PLS-PM analysis was performed via the R package (named “plspm”, version 0.4.7) to explore the direct and indirect relationships among environmental variables and the breast cancer risk. PLS-PM facilitates the analysis of high-dimensional datasets within a less structured environment by extracting the shared structures within multiple variables. Therefore, PLS-PM provides an intuitive way to validate the plausibility of the proposed breast-cancer-risk-inducing pathway involving multiple water components. Moreover, the relative contributions of different environmental variables to breast cancer were quantified by calculating the corresponding path coefficients of the PLS-PM model. All the environmental variables used in the PLS-PM model are detailed in Table S2. Among them, the DOM and EDCs were set as latent variables (i.e., the composite variable that was indicated by multiple measured variables) reflecting water quality, and consisted of a number of typical DOM characterization parameters and EDCs contaminant contents, respectively. Since some socioeconomic characteristics of the female population (FPCs) are known breast cancer risk factors, they were likewise included as latent variables reflecting human factors and for referencing the relative impacts of DOM and EDCs, consisting of typical parameters recommended by the United Nations, i.e., the life expectancy at birth, the mean years of schooling, and the disposable income per capita (Table S3). All the input data were standardized to eliminate dimension differences before the PLS-PM analysis.
3. Results and Discussion

3.1. Overall Levels of Toxicity with Breast-Cancer-Associated Properties in Drinking Water and Source Water

Figure 2a,b summarize the results of the CHO cell cytotoxicity and thiol-specific reactivity for raw and treated water samples in each DWTP. Compared to raw water, most treated water showed varying reductions in the CTI values. This is similar to many DWTPs in other regions, where the disinfection process generally reduces the cytotoxicity of raw water [41]. In contrast, the TRI values of all treated waters increased significantly compared to raw water. This was mainly because the DBPs in chlorinated treated water can be broadly captured by thiol and are dominant drivers of oxidative stress in human breast cancer cells [27].

Figure 2. The values of (a) the cytotoxicity index (CTI) and (b) the thiol-specific reactivity index (TRI) of raw and treated water from drinking water treatment plants in the 11 study zones. The error bars represent the standard error of the mean (number of water samples, n = 4). For study zones with more than one drinking water treatment plant (i.e., zones 1, 2, 3, 10, and 11), the suffixes a/b/c are used to distinguish between different drinking water treatment plants in the same zone.

Correlations between the treated water’s CTI and TRI values with the age-standardized breast cancer (c,d) incidence and (e,f) mortality. The solid lines and the enclosed areas are linear fits and the 95% confidence level with a correlation coefficient of $r$; ***, significance levels of $p < 0.05/0.01$.

Since the cytotoxicity analysis measured the reduction in CHO cell density after exposure to water samples, lower CTI values reflect a greater promotion of cell proliferation. Notably, cell proliferation through the ER-mediated pathway is considered a hallmark of estrogenic action [42], and the stimulation of proliferation due to the overexpression of ER signaling may be involved in early breast carcinogenesis [43]. Many in vitro studies have demonstrated that EDCs with estrogenic activity are key environmental factors in inducing the overexpression of ERs, which can stimulate ER-induced cell proliferation at physiologically achievable low levels [24,44]. Therefore, the EDC distributions that we previously identified in raw water from the study area may account for the variation in the CTI: compared to the low-incidence region (i.e., zone 10), the high-breast-cancer-incidence region (i.e., zone 2) had higher contents of EDCs and lower CTI values.
error bars represent the standard error of the mean (number of water samples, \( n = 4 \)). For study zones with more than one drinking water treatment plant (i.e., zones 1, 2, 3, 10, and 11), the suffixes a/b/c are used to distinguish between different drinking water treatment plants in the same zone. Correlations between the treated water’s CTI and TRI values with the age-standardized breast cancer (c,d) incidence and (e,f) mortality. The solid lines and the enclosed areas are linear fits and the 95% confidence level with a correlation coefficient of \( r; */** \), significance levels of \( p < 0.05/0.01 \).

The correlation between two toxicity indexes of treated water and the breast cancer incidence and mortality is shown in Figure 2c,d and Figure 2e,f, respectively. In treated water, the breast cancer incidence showed a significant negative correlation with the CTI (\( r = -0.99, p < 0.05 \)) and a weak positive correlation with the TRI (\( r = 0.16, p > 0.05 \)). In contrast, the breast cancer mortality presented a significant positive correlation with the TRI (\( r = 0.79, p < 0.01 \)), but no correlation with the CTI (\( r = 0.01, p > 0.05 \)). Compared to treated water, the toxicity indexes of raw water can serve as a proxy for the contaminated background value in the living environment of the local population. In raw water, the CTI maintained a strong negative correlation with the breast cancer incidence (\( r = -0.79, p > 0.05 \)), while the positive correlation between the TRI and mortality was attenuated (\( r = 0.37, p > 0.05 \)) (Figure S4). This implies an association between the toxicity (CTI and TRI) of water sources and the breast cancer risk in local populations, and this association was strengthened by a DWTP treatment.

Since the cytotoxicity analysis measured the reduction in CHO cell density after exposure to water samples, lower CTI values reflected a greater promotion of cell proliferation. Notably, cell proliferation through the ER-mediated pathway is considered a hallmark of estrogenic action [42], and the stimulation of proliferation due to the overexpression of ER signaling may be involved in early breast carcinogenesis [43]. Many in vitro studies have demonstrated that EDCs with estrogenic activity are key environmental factors in inducing the overexpression of ERs, which can stimulate ER-induced cell proliferation at physiologically achievable low levels [24,44]. Therefore, the EDC distributions that we previously identified in raw water from the study area may account for the variation in the CTI [29]: compared to the low-incidence region (i.e., zone 10), the high-breast-cancer-incidence region (i.e., zone 2) had higher contents of EDCs and lower CTI values. Due to the inefficiency of conventional DWTPs in eliminating EDCs, most EDCs can penetrate the treatment process and remain in treated water [45], which also happens in the study area. This may explain the strong correlation between the CTI and the breast cancer incidence for raw water that persists in treated water. Unlike the CTI, the correlation between the TRI and breast cancer was observed in terms of mortality, and was not significant before the treatment by DWTPs. This suggests that DWTPs have an enhanced TRI toxicity in some raw water components, and DOM mixtures as precursors to DBPs are prioritized components to be considered. Overall, the strong correlations between the breast cancer incidence and mortality with the CTI and the TRI, respectively, suggest that the induction of estrogenic effects and oxidative stress, two broad-spectrum communal properties associated with breast cancer, coexist in the mixture extracts of source and drinking water. Furthermore, these properties originate from the impact of DWTPs and source water contamination.

3.2. Identification of Key DOM with Breast-Cancer-Associated Properties in Water Mixture Contaminants

For the two toxicological commonalities widely expressed in water mixtures that are associated with breast cancer, the potential drivers remain largely unknown. Herein, the DOM, the dominant source of the organic fraction in water mixtures, is further resolved to identify possible key components. For the dataset containing all the samples, six fluorescent DOM components were determined by PARAFAC modeling (Figure 3). The robustness of the six DOM components was validated by the residual error and a split-half analysis (Figures S5 and S6) [46]. Among these DOM components, FD2, FD4, and FD5 were the components of interest in treated water, as they showed noticeable correlations with the breast cancer incidence or mortality (\( |r| > 0.40 \)) (Figure 4a,b), while the remaining
components showed a weak or no correlation (\(|r| < 0.40\)) (Figure S7). Notably, FD5 was the only component that showed high negative correlations with both the incidence \((r = -0.73, p > 0.05)\) and mortality \((r = -0.62, p < 0.05)\), suggesting that it may be a protective factor against breast cancer. In contrast, FD2 and FD4 showed a positive correlation with the breast cancer mortality \((r = 0.42-0.49, p > 0.05)\), but no correlation with the incidence. These components were detected in both the raw and treated water of the study area (Figure S8) and surface water worldwide (Table S4), indicating that they originate from water sources and are not PRB-region-specific.

To further identify the key DOM with breast-cancer-associated properties in water mixtures, a correlation analysis of the DOM content with the CTI and the TRI of treated water was carried out (Figure 4c,d). Among the components, FD2 and FD4 showed a relatively strong positive correlation with the TRI \((r = 0.54, p < 0.05)\), along with a certain degree of positive correlation with the CTI \((r = 0.24-0.39, p > 0.05)\). This suggests that FD2 and FD4 may have cytotoxic and oxidative-stress-inducing potentials. Notably, the relatively strong correlation of FD2 and FD4 with the TRI was also observed in raw water \((r = 0.42, p > 0.05)\) and \((r = 0.61, p < 0.01)\), respectively), and the correlation of FD2 and FD4 with the CTI in treated water was enhanced compared to that of raw water (Figure S9). This indicates that the risk caused by FD2 and FD4 originates from water sources, and that these components may be important precursors of toxic DBPs. The component-matching results from the OpenFlour database further revealed that both FD2 and FD4, with a high reactivity, were from anthropogenic sources (Table S4). Among them, FD2 was suggested to originate from petroleum derivatives and was strongly correlated with acute toxicity \([47]\), which agrees with our results. On the other hand, FD4, commonly found in urbanized watersheds, was associated with surface runoff, and was considered a wastewater-related component \([48–50]\).
To further identify the key DOM with breast-cancer-associated properties in water mixtures, a correlation analysis of the DOM content with the CTI and the TRI of treated water was carried out (Figure 4c, d). Among the components, FD2 and FD4 showed a relatively strong positive correlation with the TRI \( (r = 0.41, p > 0.05 \text{ and } r = 0.54, p < 0.05) \), along with a certain degree of positive correlation with the CTI \( (r = 0.24 \text{–} 0.39, p > 0.05) \). This suggests that FD2 and FD4 may have cytotoxic and oxidative-stress-inducing potentials. Notably, the relatively strong correlation of FD2 and FD4 with the TRI was also observed in raw water \( (r = 0.42, p > 0.05 \text{ and } r = 0.61, p < 0.01) \), and the correlation of FD2 and FD4 with the CTI in treated water was enhanced compared to that of raw water (Figure S9). This indicates that the risk caused by FD2 and FD4 originates from water sources, and that these components may be important precursors of toxic DBPs.

The component-matching results from the OpenFlour database further revealed that both FD2 and FD4, with a high reactivity, were from anthropogenic sources (Table S4). Among them, FD2 was suggested to originate from petroleum derivatives and was strongly correlated with acute toxicity [47], which agrees with our results. On the other hand, FD4, commonly found in urbanized watersheds, was associated with surface runoff, and was considered a wastewater-related component [48–50].

The negative correlation of FD5 and the TRI \( (r = -0.33, p > 0.05) \) implied that its protective effect is associated with antioxidant properties (Figure 4d). Unlike FD2 and FD4, FD5 is derived from autochthonous (e.g., phytoplankton production) and terrestrial (e.g., lignin-rich biomass) sources, pointing to terrestrial fluorescent material of forested catchments [51–54] (Table S4). Since the \( S_{275–295} \) of DOM (i.e., UV spectral slope at 275–295 nm) is inversely related to the total lignin content [55], it was utilized to examine the association of the total lignin content with breast cancer and water toxicity (Figure S10). The \( S_{275–295} \) in treated water was positively correlated with both the breast cancer incidence and the mortality \( (r = 0.97, p < 0.01 \text{ and } r = 0.43, p > 0.05, \text{ respectively}) \), while it was
negatively correlated with the CTI (r = −0.40, p > 0.05), indicating that the breast-cancer-protective effect of lignin-related components may also be associated with anti-proliferative properties. The FTIR spectra of DOM further confirmed the widespread nature of lignin-related components in the study area (Figure S11), with the absorption peaks referring to the O-H and C-O groups of phenols (1420 cm\(^{-1}\)) and the aromatic C-H of the guaiacyl units in lignin (1140, 860, and 810 cm\(^{-1}\)) [56–60]. Interestingly, the relative intensities of the major lignin-characterized peaks (i.e., 1140 and 860 cm\(^{-1}\)) showed negative correlations with the breast cancer incidence and mortality, as well as positive correlations with the CTI (Figure S12). This agrees with the results for the total lignin content (i.e., \(S_{275–295}\)) and illustrates the breast-cancer-associated protective properties of lignin-related components from another angle. It should be noted that the antioxidant and anti-proliferative properties of lignin-related components in breast cancer cells [61] and their effect in reducing the breast cancer risk in populations have been confirmed [62,63]; hence, this supportive evidence makes such components a protective factor of note in drinking water.

Overall, anthropogenically sourced DOM components (i.e., FD2 and FD4) are potential drivers of the breast-cancer-associated oxidative-stress-inducing properties in water mixtures. In contrast, lignin-related components with a natural source (e.g., FD5) may influence the breast cancer risk by modulating the level of oxidative stress or estrogenic effects in water mixtures through their antioxidant or anti-proliferative properties.

3.3. Identification of Key EDCs with Breast-Cancer-Associated Properties in Water Mixture Contaminants

Due to their ability to affect the mammalian endocrine system, EDCs are another suspected factor involved in the induction of breast-cancer-associated properties in water mixtures. Several EDCs associated with breast cancer, including nonylphenol and inorganic contaminants (e.g., metalloids and metals), were identified in the water sources of some sub-study zones in our previous work [29]. In this study, we further examined their overall distribution in raw and treated water. The contents of nonylphenol and inorganic contaminants are summarized in Figure 5a and Tables S5 and S6, respectively. For the components currently included as controlled contaminants in the Chinese drinking water quality standard, i.e., As (≤0.01 mg/L), Sb (≤0.005 mg/L), Ba (≤0.7 mg/L), Ni (≤0.02 mg/L), B (≤1.0 mg/L), nitrate (≤10 mg/L), and fluoride (≤1.0 mg/L), they were all below the criteria limits in the treated water of all the study zones. However, our results confirmed that these EDCs from raw water penetrate the treatment process and remain in treated water, similar to other conventional DWTPs worldwide [45].

As one of the widely suspected key EDCs for breast cancer, nonylphenol in treated water exhibits a strong positive correlation with both the breast cancer incidence (r = 0.99, p < 0.01) and the mortality (r = 0.50, p > 0.05) (Figure 5b,c). Similar correlations were likewise observed in raw water (Figure S13a,b), suggesting that, despite a certain amount of nonylphenol being reduced by conventional DWTPs, the risk induced by nonylphenol in raw water was still largely transferred to treated water. In correlation analyses with water toxicity, nonylphenol showed a significant negative correlation with the CTI (r = −0.52~−0.64, p < 0.05), but it did not show a correlation with the TRI, neither in raw water (Figure S13c,d) nor treated water (Figure 5d,e). This implies that nonylphenol is a potential key driver of the induction of intense cell proliferation in water mixtures. This is likewise supported by in vitro experimental results on individual exposure to nonylphenol, which causes aberrant cell proliferation through ER-mediated signaling [20].
likewise supported by in vitro experimental results on individual exposure to nonylphenol, which causes aberrant cell proliferation through ER-mediated signaling [20].

**Figure 5.** (a) Nonylphenol (NP) concentrations in raw and treated water from drinking water treatment plants in the 11 study zones. The total concentration of NP is the sum of the branched 4-NP and 4-n-NP. The correlation of the total NP in treated water with the age-standardized breast cancer (b) incidence and (c) mortality, and with (d) the cytotoxicity index (CTI) and (e) the thiol-specific reactivity index (TRI) values. The lines and the enclosed areas are linear fits and the 95% confidence level with a correlation coefficient of \( r \); ** significance levels at \( p < 0.01 \).

For multiple of the examined inorganic components, the suspect species were first screened by their correlation with breast cancer. Among them, Sb, Co, Ba, Ni (Figure 6a,b), B (Figure S14a,b), and nitrate (Figure S15a,b) were of interest due to their strong correlations with the breast cancer incidence and/or the mortality \((|r| > 0.40)\). Notably, B was negatively correlated with both the breast cancer incidence \((r = -0.41, p > 0.05)\) and the mortality \((r = -0.65, p < 0.05)\), indicating a potential protective factor. In contrast, the remaining contaminants in treated water were potential risk factors for breast cancer. Like nonylphenol, these suspect inorganic contaminants in raw water also exhibited similar correlation coefficients with breast cancer compared to treated water, again suggesting that their breast cancer risk originates from the impact of source water.
In brief, amines, also listed as potential carcinogens by the IARC [66]. For B, its protective effect may be explained by the anti-proliferative and antioxidant tendencies indicated by its correlations with the CTI. For Ni, the correlation coefficient was 0.60, which showed no correlation with the CTI, suggesting that Ni mainly had oxidative-stress-induced tendencies. This could be due to the fact that the induction of oxidative stress is also a non-negligible carcinogenesis pathway for Ni in mammalian cells [36].

To further identify the key EDCs with breast-cancer-associated properties in water mixtures, correlation analyses of these suspect inorganic contaminants with the CTI and the TRI were performed. For the correlation with the CTI, Sb (r = −0.34, p > 0.05), Co (r = −0.55, p < 0.05), and Ba (r = −0.28, p > 0.05) all showed a negative correlation (Figure 6c), demonstrating a tendency to promote cell proliferation. This result was supported by in vitro experiments targeting individual Sb, Co, and Ba exposure to human breast cancer cells [36]. For the TRI, significant positive correlations only occurred for Ni (r = 0.60, p < 0.05) (Figure 6d), which showed no correlation with the CTI, suggesting that Ni mainly had oxidative-stress-induced tendencies. This could be due to the fact that the induction of oxidative stress is also a non-negligible carcinogenesis pathway for Ni in mammalian cells [64]. It is worth noting that Sb (group 2A/3), Co (group 2A), and Ni (group 1/2B) have all been listed as potential carcinogens by the International Agency for Research on Cancer (IARC); hence, their risks in drinking water require further attention. On the other hand, although nitrate showed a correlation with the TRI (Figure S15c,d), no relevant toxicological evidence has been found. As an indicator of wastewater discharge, the epidemiologic evidence of the association of nitrate in water with breast cancer was inconsistent [10,65,66]. However, more explicit evidence of the association between the nitrate intake and breast cancer arises from its human metabolites, N-nitroso compounds, also listed as potential carcinogens by the IARC [66]. For B, its protective effect may be explained by the anti-proliferative and antioxidant tendencies indicated by its correlations with the CTI.

Figure 6. The correlation between the concentrations of metals/metalloids of interest (Sb, Co, Ba, and Ni) with the age-standardized breast cancer (a) incidence and (b) mortality, and with (c) the cytotoxicity index (CTI) and (d) the thiol-specific reactivity index (TRI) values in treated water. The lines and the enclosed areas are linear fits and the 95% confidence level with a correlation coefficient of r; */**, significance levels at p < 0.05/0.01.
with the CTI ($r = 0.18, p > 0.05$) and the TRI ($r = -0.32, p > 0.05$) (Figure S14c,d), which were supported by other toxicologic and epidemiologic evidence of breast cancer [38,67,68]. In brief, nonylphenol, Sb, Co, and Ba are potential drivers of the breast-cancer-associated estrogenic effect properties in water mixtures, whereas Ni is another possible driver of the induction of oxidative stress properties.

3.4. Breast-Cancer-Risk-Inducing Paths Based on Water Toxicological Commonalities

The preceding analysis between water contaminants, water toxicity, and the breast cancer risk identified the breast cancer-associated toxicological commonalities in water mixtures and the potential drivers. However, considering that the effects of some contaminants may be masked in the water mixture, the overall impacts of all the examined DOM and EDCs on water toxicity and breast cancer need to be further evaluated from a general perspective. Moreover, the plausibility that water contaminants influence breast cancer by altering these two toxicological commonalities in treated water has not been verified. To address these issues, a hypothesis-driven PLS-PM model based on water toxicological commonalities was constructed to validate the proposed breast-cancer-risk-inducing paths and to quantify the relative contributions of DOM and EDCs. By identifying common structures among multi-variables, PLS-PM can analyze high-dimensional water contaminant datasets in a low-structured environment, thereby characterizing common features of all the examined water contaminants.

Since the associations of water contaminants with the breast cancer incidence and mortality were not always the same, the incidence and mortality were applied as different endpoints for the PLS-PM model, and the results are shown in Figure 7a,b. The PLS-PM model’s goodness of fit (GoF in the 0–1 range) allows for a quantitative assessment of how closely the water contaminant dataset is associated with the proposed paths, with a higher GoF indicating a more reasonable path estimation. The GoFs of the PLS-PM model targeting the incidence and mortality were all greater than 0.5, indicating that the toxicological-commonality-based PLS-PM model had a relatively good interpretability for breast cancer risk. Therefore, this supports the likelihood of the proposed risk-inducing path for breast cancer on a statistical level.

For the model of breast cancer incidence (Figure 7a), the path coefficients of the CTI and the TRI for the breast cancer incidence were $-0.40 (p > 0.05)$ and $-0.10 (p > 0.05)$, respectively. This implies that estrogenic effects promoting cell proliferation dominate the incidence risk, while the role of oxidative stress promoting thiol depletion is marginal. The path coefficients of DOM and EDCs on the CTI ($0.02$ and $-0.87, p > 0.05$, respectively) further suggest that EDCs are the main drivers of estrogenic effects related to the breast cancer incidence, while the overall impact of DOM is negligible. On the other hand, for the model of breast cancer mortality (Figure 7b), the path coefficients of the CTI and the TRI for the breast cancer mortality were $-0.39 (p > 0.05)$ and $0.95 (p < 0.05)$, respectively. This suggests that both the estrogenic effects and oxidative-stress-induced effects of the water mixtures are associated with the mortality risk of breast cancer. The path coefficients of DOM and EDCs on the CTI were $0.73 (p < 0.001)$ and $-0.48 (p < 0.05)$, respectively. This implies that certain types of DOM (e.g., lignin-related components) may antagonize a certain degree of the estrogenic effects related to the breast cancer mortality, which are mainly induced by EDCs. The path coefficients of DOM and EDCs on the TRI were $0.66 (p < 0.01)$ and $0.13 (p > 0.05)$, respectively. This suggests that DOM is an important factor involved in inducing the oxidative stress effects in water mixtures, while the impact of EDCs is weak.

According to the standardized total effects of the PLS-PM model, the total impact of DOM and EDCs on the breast cancer incidence was $-0.07$ and $0.31$, respectively (Figure 7c), while the impact on mortality was $0.35$ and $0.31$, respectively (Figure 7d). This further suggests that the breast cancer mortality risk induced by water mixtures was driven by the joint effect of DOM and EDCs, with a similar magnitude of impact for both. In contrast, the incidence risks were dominated by EDCs. For reference, the total effect of the FPCs,
the known risk factors related to the female population, on the breast cancer incidence and mortality was 0.60 and 0.12, revealing that the risks of DOM and EDCs in treated water mixtures are not negligible.

Figure 7. Partial least-squares path model (PLS-PM) describing the direct and indirect effects of the treated water quality and female population factors on the breast cancer age-standardized (a) incidence and (b) mortality. The composite variables DOM, EDCs, and female population characteristics (FPCs) are explained in Table S2. The loadings of the factors contained in the composite variable are shown in Figure S16. Larger path coefficients are displayed as thicker arrows, with red and blue indicating positive and negative effects, respectively. The value on the arrow indicates the path coefficient; */**/***, significance level at \( p < 0.05/0.01/0.001 \). The standardized total effects on the breast cancer risk, cytotoxicity index (CTI) value, and thiol-specific reactivity index (TRI) value were derived from the PLS-PM model targeting the (c) incidence and (d) mortality.
Overall, the breast cancer incidence is primarily associated with the stronger estrogenic properties of the water mixtures (i.e., a lower CTI), which were mainly induced by EDCs. In contrast, the breast cancer mortality is associated with both a stronger oxidative-stress-inducing potential and estrogenic properties (i.e., a higher TRI and a lower CTI) resulting from the combined impacts of EDCs and DOM. While conventional DWTPs can remove a small number of EDCs, they also unavoidably lead to an enhanced toxicity (e.g., TRI) by some water components during chlorination; thus, the risk of breast cancer in treated water is an interaction outcome of these two processes and reflects the dilemma of conventional DWTPs.

4. Conclusions

By utilizing the latest epidemiology data and an exhaustive water toxicological and chemical characterization, coupled with PLS-PM statistics modeling, our findings indicate that the induction of estrogenic effects and oxidative stress are two broad-spectrum communal properties that are associated with breast cancer for mixed extracts of source and drinking water. The water-mixture-promoted mammalian cell proliferation via estrogen receptors (i.e., estrogenic effect) and the greater thiol depletion (i.e., oxidative stress) are two important paths for the drinking-water-toxicity-induced breast cancer risk, mainly driven by EDCs and DOM, respectively. By combining the existing knowledge of the breast cancer environmental etiology, key EDCs and DOM species with these two breast-cancer-associated properties were identified among the water mixture, including nonylphenol, Sb, Co, Ba, Ni, and anthropogenically sourced DOM components (FD2 and FD4). Moreover, a lignin-related DOM component (FD5) was a potential protective factor. The PLS-PM model further confirmed that EDCs are a forcing factor for the incidence risk, while the mortality risk is a joint effect of EDCs and DOM. It should be noted that, although our study provides multi-faceted evidence of associations between drinking water and breast cancer, it is still insufficient to indicate a clear causal relationship. Given that breast cancer is an outcome caused by multiple factors, future studies could look at the overall toxicity of drinking water mixtures to comprehensively screen for possible environmental factors. Overall, combined exposures from contaminants with estrogenic effects and oxidative-stress-inducing potential in drinking water are a non-negligible source of breast cancer risk.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/w16050702/s1.

Author Contributions: Conceptualization, S.P., C.G. and X.C.; Methodology, S.D.; Formal analysis, S.P.; Investigation, S.P., H.D., Y.Z. and C.Y.; Resources, S.D. and C.G.; Writing—original draft, S.P.; Supervision, X.C.; Funding acquisition, S.D., C.G. and X.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was financially supported by the National Natural Science Foundation of China: 51861125203, 52370020, 52000184.

Data Availability Statement: The data are contained within the article and Supplementary Materials.

Conflicts of Interest: The authors declare no conflicts of interest.

References


60. Broder, T.; Blodau, C.; Biester, H.; Knorr, K.-H. Peat decomposition records in three pristine ombrotrophic bogs in southern Patagonia. *Biogeosciences* 2012, 9, 1479–1491. [CrossRef]


Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.