ORIGINAL RESEARCH

Estimation of Personal Symptom Networks Using the Ising Model for Adult Survivors of Childhood Cancer: A Simulation Study with Real-World Data Application

Yiwang Zhou [1], Madeline R Horan[2], Samira Deshpande[1], Kirsten K Ness[2], Melissa M Hudson[3], I-Chan Huang[2,]*, Deokumar Srivastava[1,]*

[1]Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, TN, USA; [2]Department of Epidemiology and Cancer Control, St. Jude Children's Research Hospital, Memphis, TN, USA; [3]Department of Oncology, St. Jude Children's Research Hospital, Memphis, TN, USA

*These authors contributed equally to this work

Correspondence: Yiwang Zhou, Department of Biostatistics, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Mail Stop 768, Memphis, Tennessee, TN 38105, USA, Tel +1 901-595-6736, Email yiwang.zhou@stjude.org

Purpose: Childhood cancer survivors experience interconnected symptoms, patterns of which can be elucidated by network analysis. However, current symptom networks are constructed based on the average survivors without considering individual heterogeneities. We propose to evaluate personal symptom network estimation using the Ising model with covariates through simulations and estimate personal symptom network for adult childhood cancer survivors.

Patients and Methods: We adopted the Ising model with covariates to construct networks by employing logistic regressions for estimating associations between binary symptoms. Simulation experiments assessed the robustness of this method in constructing personal symptom network. Real-world data illustration included 1708 adult childhood cancer survivors from the St. Jude Lifetime Cohort Study (SJLIFE), a retrospective cohort study with prospective follow-up to characterize the etiology and late effects for childhood cancer survivors. Patients' baseline symptoms in 10 domains (cardiac, pulmonary, sensation, nausea, movement, pain, memory, fatigue, anxiety, depression) and individual characteristics (age, sex, race/ethnicity, attained education, personal income, and marital status) were self-reported using survey. Treatment variables (any chemo or radiation therapy) were obtained from medical records. Personal symptom network of 10 domains was estimated using the Ising model, incorporating individual characteristics and treatment data.

Results: Simulations confirmed the robustness of the Ising model with covariates in constructing personal symptom networks. Real-world data analysis identified age, sex, race/ethnicity, education, marital status, and treatment (any chemo and radiation therapy) as major factors influencing symptom co-occurrence. Older childhood cancer survivors showed stronger cardiac-fatigue associations. Survivors of racial/ethnic minorities had stronger pain-fatigue associations. Female survivors with above-college education demonstrated stronger pain-anxiety associations. Unmarried survivors who received radiation had stronger association between movement and memory problems.

Conclusion: The Ising model with covariates accurately estimates personal symptom networks. Individual heterogeneities exist in symptom co-occurrence patterns for childhood cancer survivors. The estimated personal symptom network offers insights into interconnected symptom experiences.

Keywords: bootstrap testing, individual heterogeneity, Ising model, network analysis, sociodemographic

Introduction

Survivors of pediatric cancer often report that they experienced physical, somatic, or psychological symptoms during therapy. Symptoms may persist post-therapy or emerge years after completion of therapy. A meta-analysis based on

461

114,000 children, adolescents, and young adults with cancer suggests that symptoms like fatigue, pain, and anxiety are prevalent.[1] Based on a 37-item battery, a recent study found that 50% of the long-term adult survivors of childhood cancer experienced a moderate to high burden of multi-symptoms aggregated in physical, somatization, and/or psychologic domains.[2] This high symptom burden was significantly associated with clinically-assessed medical conditions and poor quality of life.[2,3]

Although previous studies reported the prevalence of varying symptoms among individuals with cancer, these studies merely provide a snapshot at the level of individual symptoms or aggregated clusters of different symptoms. Patterns of concurrent symptoms may exist, including different 1) presentations of the type and number of multi-symptoms and 2) degrees of association between concurrent symptoms. For example, a survivor who experiences headache may experience other symptoms simultaneously (eg, fatigue and depression). The associations between headache and the other three symptoms may be stronger than the associations among the other three individual symptoms themselves. In this case, headache can be considered as a central symptom. Application of network sciences can shed light on the intricate patterns of multiple symptoms in cancer survivors, offering valuable information to enhance clinical decision-making processes.[4]

Successful statistical methods for analyzing network structures have emerged over the past decades.[5–9] However, current symptom network analysis is often performed based on the average individual without considering individual heterogeneities.[4,10–18] While extended statistical methods have been proposed to incorporate individual or group factors[19–21] in the estimation of symptom networks, to the best of our knowledge, there has been no research conducted to comprehensively evaluate how individual heterogeneities introduced by sociodemographic and/or clinical characteristics influence associations between multi-symptoms for cancer patients and survivors. Such analysis can help elucidate the etiology, progression, and consequence of symptom burden, and facilitate the design of precision-based symptom intervention. Here, the term "personal symptom network" pertains to understanding the impact of individual characteristics on the associations between symptoms within a network. This elucidation will provide a distinct structure of a symptom network determined by individual factors. It is important to emphasize that this concept draws upon the notable individualized treatment rules in precision medicine, where personalized decision functions are often derived from cross-sectional data based on individual characteristics.[22] This approach differs from the between-subject network estimated using time-series data.[23–25]

While recent research has proposed the use of Ising model with covariates to compare group differences of networks,[19–21] a comprehensive evaluation of this model in estimating personal symptom networks by considering a broader spectrum of individual factors, requires extensive simulation experiments. Hence, the first objective of this study is to assess the robustness and stability of the Ising model with covariates in constructing personal symptom network through simulations. The second objective was to use the real-world data to estimate the personal symptom network for adult survivors of childhood cancer from the St. Jude Lifetime Cohort Study (SJLIFE).[2] We aimed to construct the network on patients' baseline symptoms across 10 domains (cardiac symptoms, pulmonary symptoms, sensation abnormality, nausea, movement problems, pain, memory problems, fatigue, anxiety, and depression). By incorporating individual characteristics such as age, sex, race/ethnicity, attained education, personal income, and marital status, as well as treatment data (ie, ever received chemotherapy or radiation), we utilized the Ising model to estimate the covariate-related pairwise associations between symptoms. The Ising model[26–28] was chosen for network estimation because we focused on the presence of symptoms (binary indicators) rather than the severity of symptoms (continuous indicators). After network estimation, the accuracy and stability of the personal symptom network was evaluated using bootstrap testing.[29] Information derived from individualized network modeling will facilitate identifying the individual characteristics that influence associations between symptoms and central symptoms that played a role in triggering the overall network structure among adult survivors of childhood cancer.

The remainder of the paper is organized as follows: The Methods section outlines the basics of network estimation, introduces the previously proposed Ising model with covariates, and describes the bootstrap testing procedures used to assess the accuracy and stability of the estimated network. In the Results section, we evaluate the Ising model with covariates through extensive simulations and present the estimated personal symptom network for adult survivors of childhood cancer from SJLIFE. Discussion and concluding remarks are provided in their respective sections. Details of

the algorithm of the Ising model with covariates, details of the bootstrap testing procedure, and additional simulation and real-world data analysis results are included in the Supplementary Materials.

Methods

Symptom Network Estimation

Network analysis is a graph theory-based technique used to study associations between objects.[30,31] A network contains two fundamental components: nodes and edges. In this work, nodes are binary symptoms, denoted by $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_p)$ with $\boldsymbol{Y} \in \mathcal{Y} = \{0, 1\}$. Edges are associations between two symptoms. Two symptoms are conditionally independent of each other if there is no edge between them; otherwise, they are conditionally dependent. The weight of an edge represents the strength of conditional dependency between symptoms. The structure of a network encodes certain conditional independence assumptions among symptoms. The aim of network analysis is to estimate the underlying graph from n independent and identically distributed samples $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{ip}), i = 1, \ldots, n$.

Ising model is commonly used to construct networks for binary data. Previous studies[26,27,32–34] have demonstrated that estimation of a network using the Ising model can be achieved through a total of p logistic regressions, each taking one symptom as the outcome and the other symptoms as covariates. Denoting $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)$ as the intercepts and $\boldsymbol{\Sigma} = \{\sigma_{jj'}\}, j, j' = 1, \ldots, p, j \neq j'$ as the coefficients in the logistic regressions, the full log-likelihood of the Ising model is,

$$l^*(\boldsymbol{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto \left\{ \sum_{j=1}^{p} \mu_j y_j + \sum_{j=1}^{p-1} \sum_{j'=j+1}^{p} \sigma_{jj'} y_j y_{j'} \right\} - \sum_{j=1}^{p} \log\left(1 + \exp\left(\mu_j + \sum_{j \neq j'} \sigma_{jj'} y_{j'} \right) \right) \tag{1}$$

where μ_j represents the threshold of symptom y_j (ie, the probability of y_j taking value 1) and $\sigma_{jj'}$ represents the pairwise association (ie, the conditional dependence) between symptoms y_j and $y_{j'}$. If $\sigma_{jj'} = 0$, y_j and $y_{j'}$ are independent of each other given all the other variables; otherwise, y_j and $y_{j'}$ are conditionally dependent. To further balance the complexity of the network with the information available from the data, Lasso[35] is applied to impose a l_1-penalty on the parameters. Then the optimization problem for each logistic regression becomes

$$\min_{\mu_j, \sigma_j^{(j)}} l^*\left(y_j | \mu_j, \sigma_j^{(j)} \right), \text{subject to} \sum_{j \neq j'} \left| \sigma_{jj'} \right| \leq \rho \tag{2}$$

where $\sigma_j^{(j)}$ denote the column j of matrix $\boldsymbol{\Sigma}$. This effectively shrinks negligible effects to zero and creates a sparse network. Different tuning parameters ρ will result in different structures estimated by Lasso. The eLasso procedure selects the optimal structure according to the minimization of an extended Bayesian information criterion (eBIC),[26,36,37]

$$\text{eBIC} = -2l^* + |J| \log(n) + 2\eta |J| \log(p - 1)$$

where $|J|$ is the number of connected nodes selected by Lasso at a specific ρ and η is the hyperparameter of eBIC.

Personal Symptom Network Estimation

In this study, the personal symptom network was constructed using the Ising model with covariates.[19–21] With the personal symptom network, we can identify personal factors that significantly influence the association between certain symptoms. Incorporating individual characteristics into the Ising model is achieved by expanding $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.[19] Denote $\boldsymbol{X} = (X_1, \ldots, X_d) \in \mathbb{R}^d$ as a d-dimensional vector of individual features. The Ising model with covariates includes a linear sum of $\boldsymbol{X}$ into the thresholds and association parameters, making the log-likelihood as

$$l^*(\boldsymbol{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto \left\{ \sum_{j=1}^{p} \left(\mu_j + \sum_{l=1}^{d} \alpha_{jl} x_l \right) y_j + \sum_{j=1}^{p-1} \sum_{j'=j+1}^{p} \left(\sigma_{jj'} + \sum_{l=1}^{d} \gamma_{jj'l} x_l \right) y_j y_{j'} \right\}$$
$$- \sum_{j=1}^{p} \log\left(1 + \exp\left(\mu_j + \sum_{l=1}^{d} \alpha_{jl} x_l + \sum_{j' \neq j} \left(\sigma_{jj'} + \sum_{l=1}^{d} \gamma_{jj'l} x_l \right) y_{j'} \right) \right) \tag{3}$$

The threshold for y_j becomes $\mu_j + \sum_{l=1}^{d} \alpha_{jl} x_l$ and the pairwise association between y_j and $y_{j'}$ becomes $\sigma_{jj'} + \sum_{l=1}^{d} \gamma_{jj'l} x_l$, both of which include a linear sum of x_l, making the thresholds and pairwise associations dependent on individual

characteristics. Such expansion actually adds the major effects of x (ie, $\sum_{l=1}^{d} \alpha_{jl} x_l$) and interaction terms of xy (ie, $\sum_{l=1}^{d} \gamma_{jj'l} x_l y_{j'}$) into logistic regression. Estimation of parameters $\boldsymbol{\mu}, \boldsymbol{\Sigma}, \alpha$ and γ can be made based on this extended logistic regression model. Parameter μ_j is the threshold of y_j when all $x_l = 0$, and parameter α_{jl} is the difference of the threshold of y_j when x_l changes from 0 to 1 adjusting for all the other covariates. Based on α_{jl}, we can evaluate the influence of x_l on the threshold of y_j. Similarly, parameter $\sigma_{jj'}$ is the strength of pairwise association between y_j and $y_{j'}$ when all $x_l = 0$, and parameter $\gamma_{jj'l}$ is the difference of the strength of pairwise association between y_j and $y_{j'}$ when x_l changes from 0 to 1 adjusting for all the other covariates. Based on $\gamma_{jj'l}$ we can examine the effect of x_l on the conditional dependency between y_j and $y_{j'}$. With the estimates $\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}, \hat{\alpha}$ and $\hat{\gamma}$, we can construct personal symptom networks. Pseudo-codes for estimating personal symptom networks are provided in [Supplementary Algorithm 1](#).

Evaluation of Network Accuracy

After personal symptom network estimation, we evaluated the accuracy and stability of the constructed network based on bootstrap testing following Epskamp et al[29] using data splitting inference.[38] Specifically, we randomly split the dataset into two parts, using one portion for network estimation (ie, model selection) and the other for bootstrap based on the selected model. The bootstrap routines include 1) assessing the accuracy of the estimated coefficients, 2) investigating the stability of node centralities, and 3) testing differences between the estimated coefficients or centralities. A detailed description of the bootstrap testing procedure is provided in the [Supplementary Subsection: Evaluation of network accuracy](#).

Real-World Data Analysis

We applied the Ising model with covariates to analyze symptom data from adult survivors of childhood cancer who participated in the SJLIFE study. SJLIFE is a retrospective cohort study with prospective clinical follow-up and ongoing accrual, initiated in 2007, to characterize the etiology and late effects for childhood cancer survivors.[39] Eligible individuals were diagnosed and treated for pediatric cancer at St. Jude Children's Research Hospital (SJCRH) between 1962 and 2012 and survived 5 or more years after diagnosis. Cohort participation involved a comprehensive clinical, survey, and laboratory assessment at SJCRH after study enrollment. This study was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki. All procedures involving human participants were performed with the approval by the SJCRH's Institutional Review Board and informed consent was obtained from all participants.

In this work, we focused on a random sample of 2,000 adult childhood cancer survivors who participated in the SJLIFE study between 2007 and 2020 and completed a survey consisting of current symptom prevalence and individual socio-demographic characteristics at the cohort baseline. Treatment data, including any chemo or radiation therapy, were obtained from the medical records. As of 2020, SJLIFE has enrolled over 6,000 active participants.[39] Each subject self-reported the current presence of symptoms using a 37-item survey, which was further classified into 10 clinically meaningful domains: cardiac symptoms, pulmonary symptoms, sensation abnormality, nausea, movement problems, pain, memory problems, fatigue, anxiety, and depression.[2] Symptom domains were considered present if any symptom in the overall domain was endorsed (coded = 1); otherwise, it was considered absent (coded = 0). Six sociodemographic factors, including age at survey (years), sex (male; female), race/ethnicity (non-Hispanic White; other), attained education (below college/post-graduate level; other), personal income (<\$20,000; ≥\$20,000), and marital status (yes; no), as well as two treatment variables, including ever received chemotherapy (no; yes) and radiation (no; yes), were considered to estimate the covariate-related pairwise associations between symptoms. All individual characteristics were coded as binary variables, except for age at survey, which was a numeric variable scaled to mean of 0 and standard deviation (SD) of 1. The exclusion criterion was the missing of data for any symptom domain or covariates. After excluding survivors with missing information, the analysis included 1,708 survivors. We applied the Ising model with covariates to estimate the personal symptom networks by setting the tuning parameter of eLasso as $\eta = 0.25$. To gauge network accuracy, we conducted data splitting inference. Recognizing the instability of a single replication, we repeated the process 100 times. Logistic regression coefficients were then averaged across these replications to summarize symptom domain associations. Empirical coefficient distributions were derived from bootstrap values spanning the 100 replications.

Results

Simulation Results

We conducted extensive simulations to comprehensively evaluate the performance of the Ising model with covariates in estimating personal symptom networks. Additional simulation results are included in the Supplementary Subsection: Additional simulations and Supplementary Table 1. Functions for pseudo-network generation, estimation, and inference are available on GitHub at https://github.com/SamiraDesh/IndivNA.git.

In the first simulation, we examined whether the Ising model with covariates could correctly identify the true edges in the simulated personal symptom network. We simulated networks of 10 nodes $Y = (Y_1, \ldots, Y_{10}) \in \{0,1\}^{10}$. The structure of the adjacent matrix Σ was generated using the Watts-Strogatz model[29,40] with the number of neighbors set as 2 and the rewiring probability $\theta = 0.3$, which resulted in a small-world network that mimicked the networks observed in real practice. To make the network structure personalized, we assumed that some of the pairwise associations in Σ were determined by individual characteristics. We generated a total of five covariates $X = (X_1, \ldots, X_5)$, where $X_l \sim \text{Bernoulli}(0.5), l = 1 \ldots, 5$. Five random edges in Σ were selected to involve the effects of X, with the associations determined by $\gamma_{jj'l} X_l = 2X_l, l = 1, \ldots, 5$. The other associations in Σ were set as $\sigma_{jj'} = 1$. Thresholds were set as $\mu_j = -2$ and $\alpha_l = 0$. We varied the sample size as $n = \{500, 1000, 2500, 4000, 5000\}$. The tuning parameter in eBIC was set as $\eta = 0.25, 0.5, 1.0$. Personal symptom networks were estimated using the Ising model with covariates on the simulated datasets. Summary statistics of the estimation results were drawn from 200 replicates. Simulation results of the true positive rate (TPR), false positive rate (FPR), and Matthew's correlation coefficient (MCC)[10] in detecting the major and covariate-related pairwise association between nodes are summarized in Figure 1. Results showed that the TPR and MCC were high and increased with increasing sample size, while the FPR remained low and stable. The larger the value of the eBIC parameter η, the more conservative the result will be, with decreased TPR and MCC values.

In the second simulation, we assessed centrality accuracy by correlating centralities obtained from the original inference dataset's network with those derived from case-dropping bootstrap. We simulated the personal symptom network using the same settings as in simulation one. Because data splitting inference was applied for bootstrap, we increased the sample size to $n = \{1000, 2000, 5000, 8000, 10000\}$, splitting the sample in half for network estimation and bootstrap testing. Centralities of nodes in these simulated personal symptom networks were stable, since the individual variations in edge connections and node positions resulted in unique centrality calculations for each node, as the pairwise associations between nodes were affected by individual characteristics and the Watts-Strogatz model's rewiring probability $\theta = 0.3$. In contrast, we also simulated some symptom networks with unstable centralities by making all pairwise associations equal to 1 without considering individual effects and setting $\theta = 0$, which would result in chain graphs where all centralities were equal for different nodes.[29,40] The influence of stable and unstable network structures on the correlation of centralities during case-dropping bootstrap can be found in the Supplementary Materials. All these networks were constructed with parameter $\eta = 0.25$. We calculated the Spearman's rank correlations for centralities by bootstrapping 500 times along with different sampling proportions $\Delta = \{0.9, 0.8, 0.7, 0.6, 0.5\}$ and summarized the values
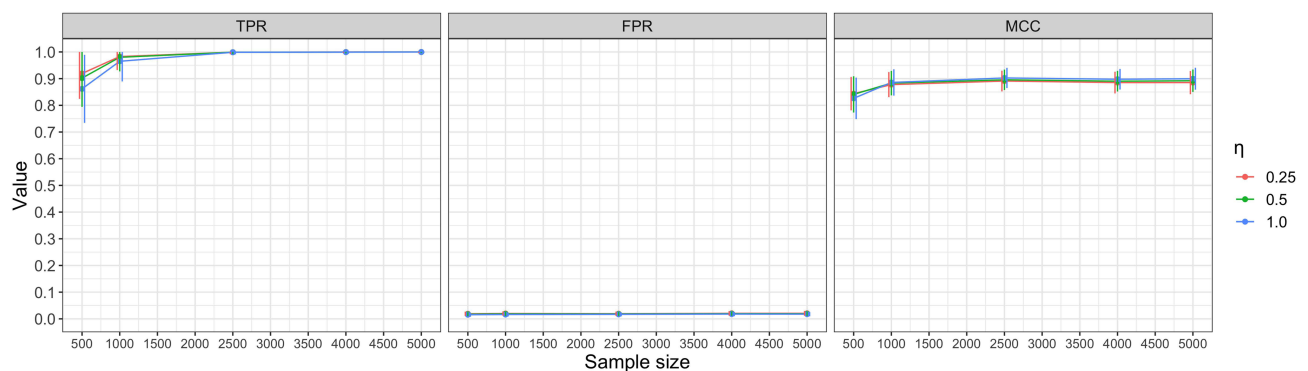

Figure 1 Identification of coefficients in the simulated personal symptom networks.
Abbreviations: TPR, true positive rate; FPR, false positive rate; MCC, Matthew's correlation coefficient.

based on 200 replicates. Figure 2 shows that, as expected, the correlations of centralities decrease with larger proportion of case-dropping. However, the correlations for strength, betweenness, and closeness remained high even with 50% dropping of samples for those networks with stable centralities (upper panel of Figure 2). In contrast, the correlations were small (below 0.20) for those networks whose centralities were not stable (lower panel of Figure 2), even if we only dropped 10% of the data. This suggests that the correlation of centralities in case-dropping bootstrap can reliably assess the accuracy of centrality indices in an estimated network.

In the third simulation, we tested differences between the estimated coefficients or centralities in the constructed personal symptom network with unstable centralities. The same settings as above were used to generate these networks. Since the rewiring probability $\theta = 0$ and all the pairwise associations equaled to 1, no coefficient or centrality should be significantly different from one another. A total of 500 bootstrap tests were performed to create the empirical distribution of the estimated coefficients and centralities. To test whether two coefficients or centralities were significantly different from each other, the empirical distribution of their differences were generated. A test was performed by checking whether zero laid in the 95% confidence interval (CI) of the empirical distribution of the differences. Summaries of the type I errors in testing the coefficients and centralities are illustrated in Figure 3. Results indicate that, with a sample size of 5000, testing coefficients can be controlled at the desired 0.05 level for type I errors. Centralities consistently maintain a controlled 0.05 type I error across varied sample sizes. Notably, the η parameter has minimal impact on type I error control.

In addition to our main findings, we performed extra simulation experiments to evaluate our method's performance in estimating personal symptom networks across different scenarios. The Supplementary Subsection: Additional simulations display these results, which affirm the method's robustness. It consistently delivered high TPR and MCC, effectively managing the FPR, except for scenarios with extreme covariate distributions, where performance slightly decreased.
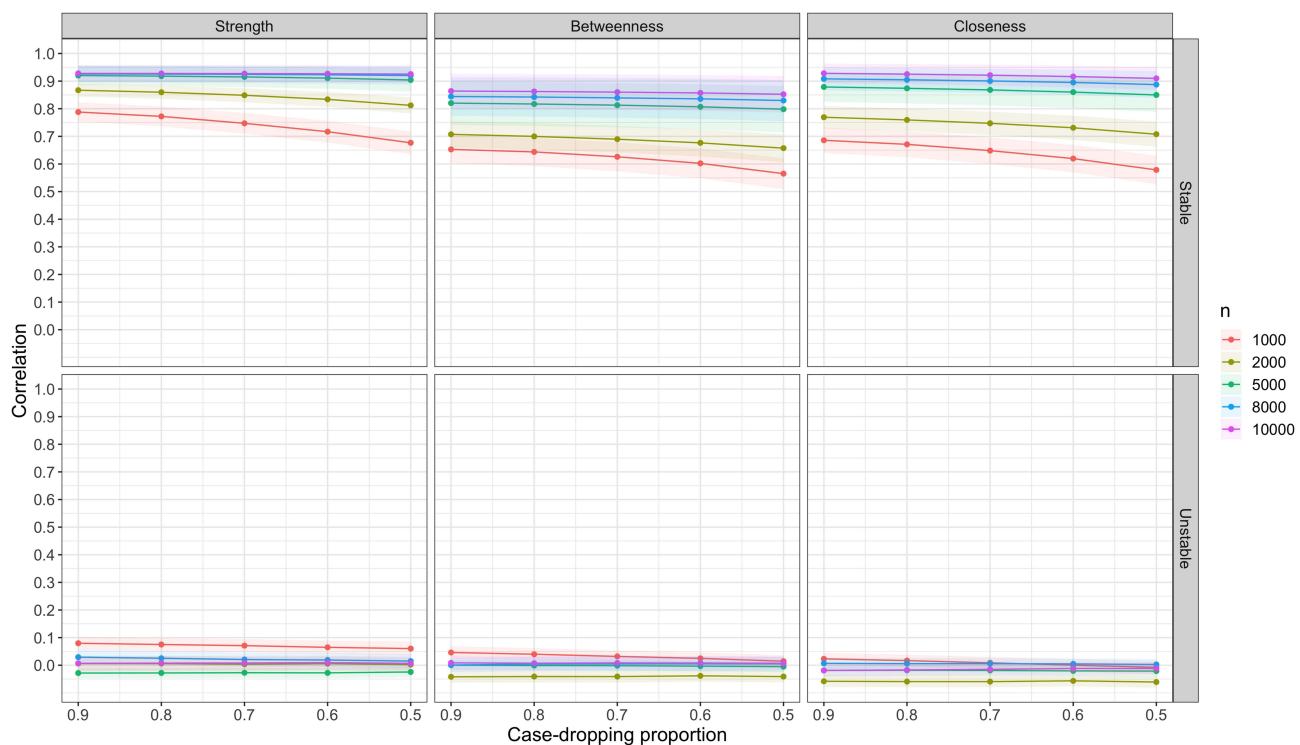
Figure 2 Correlation of centrality metrics (strength, betweenness, and closeness) between the original sample and subsets with different case-dropping proportions. The definition of the stable and unstable network structure can be found in the Supplementary Materials.
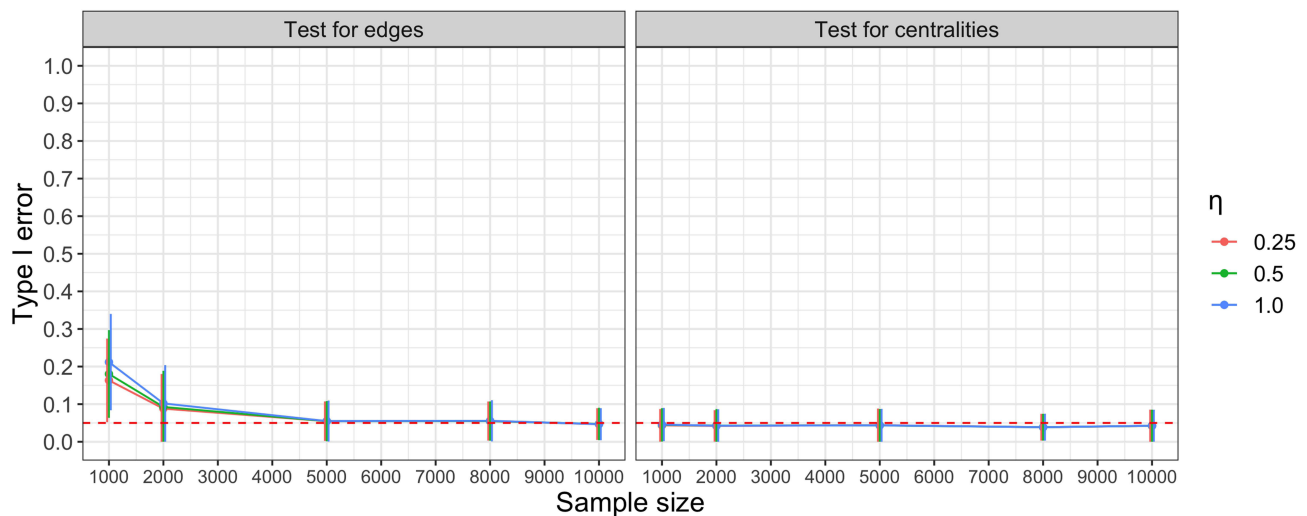
Figure 3 Type I error of testing the differences of edges and centralities for the estimated personal symptom networks.

Results of the Real-World Data Analysis

Table 1 presents the sociodemographic factors, cancer diagnoses, treatments, and prevalence of 10 symptom domains among the 1,708 adult survivors of childhood cancer included in the analysis. The survivors had a mean age of 30.9 (standard deviation (SD) 8.6) years at the time of the survey evaluation. The majority were male (52.6%) and non-Hispanic White (79.6%). Most survivors had an educational level below college/post-graduate (65.7%) and earned less than $20,000 annually (56.7%). Over half of the survivors were ever-married or had lived as married at the baseline SJLIFE visit. The two major cancer diagnoses were leukemia (35.2%) and solid tumors (31.7%), and most survivors had received chemo (84.5%) and/or radiation therapy (56.3%). The percentages of survivors experiencing the 10 symptom

Table 1 Sociodemographic Characteristics, Cancer Diagnoses, Treatments, and the Presence of 10 Symptom Domains Among 1,708 Adult Survivors of Childhood Cancer Participating in the SJLIFE Study Between 2007 and 2020. Mean (Standard Deviation (SD)) Was Reported for the Numeric Variable and Number (%) Was Reported for Categorical Variables

| Variable		Value
Individual characteristics		
Age at survey (years)		30.9 (8.6)
Sex	Male	898 (52.6%)
	Female	810 (47.4%)
Race/ethnicity	Non-Hispanic White	1360 (79.6%)
	Other	348 (20.4%)
Attained education	Below college/post-graduate level	1123 (65.7%)
	Above college/post-graduate level	585 (34.3%)
Personal income	< $20,000	969 (56.7%)
	≥ $20,000	739 (43.3%)
Marital status	Ever married or lived as married	993 (58.1%)
	Never married or lived as married	715 (41.9%)

(Continued)

Table 1 (Continued).

Variable		Value
Cancer diagnosis and treatments		
Cancer diagnosis	Leukemia	602 (35.2%)
	Lymphoma	328 (19.2%)
	Central nervous system (CNS) tumor	222 (13.0%)
	Solid tumor	541 (31.7%)
	Non-malignancy and histiocytosis	15 (0.9%)
Chemotherapy	No	264 (15.5%)
	Yes	1444 (84.5%)
Radiation	No	747 (43.7%)
	Yes	961 (56.3%)
Symptoms		
Cardiac symptoms	Yes	249 (14.6%)
	No	1459 (85.4%)
Pulmonary symptoms	Yes	176 (10.3%)
	No	1532 (89.7%)
Sensation abnormality	Yes	604 (35.4%)
	No	1104 (64.6%)
Nausea	Yes	257 (15.0%)
	No	1451 (85.0%)
Movement problems	Yes	291 (17.0%)
	No	1417 (83.0%)
Pain	Yes	449 (26.3%)
	No	1259 (73.7%)
Memory problems	Yes	470 (27.5%)
	No	1238 (72.5%)
Fatigue	Yes	313 (18.3%)
	No	1395 (81.7%)
Anxiety	Yes	528 (30.9%)
	No	1180 (69.1%)
Depression	Yes	510 (29.9%)
	No	1198 (70.1%)

domains ranged from 64.6% to 89.7%, with sensation abnormality being the least prevalent and pulmonary symptoms being the most prevalent, respectively. The associations between cancer diagnosis and the experiences of symptom domains were evaluated and reported in Supplementary Table 2.

The Ising model for personal symptom network estimation included all individual characteristics and treatment variables. Cancer diagnosis was not included because cancer treatment rather than cancer diagnosis has a more direct influence on late effects, including symptom burden, among childhood cancer survivors.[2,3] Figure 4 lists the identified associations between symptom domains and their average point estimates with 95% confidence intervals (CIs). The strongest associations were identified between symptoms independent of individual characteristics (eg, between anxiety and depression (2.20 95% CI 1.74–2.70)). The highest mean edge weight in the covariate-related associations was observed between movement and memory problems, influenced by radiation therapy (0.76, 95% CI 0–1.83), with survivors who ever received radiation therapy showing
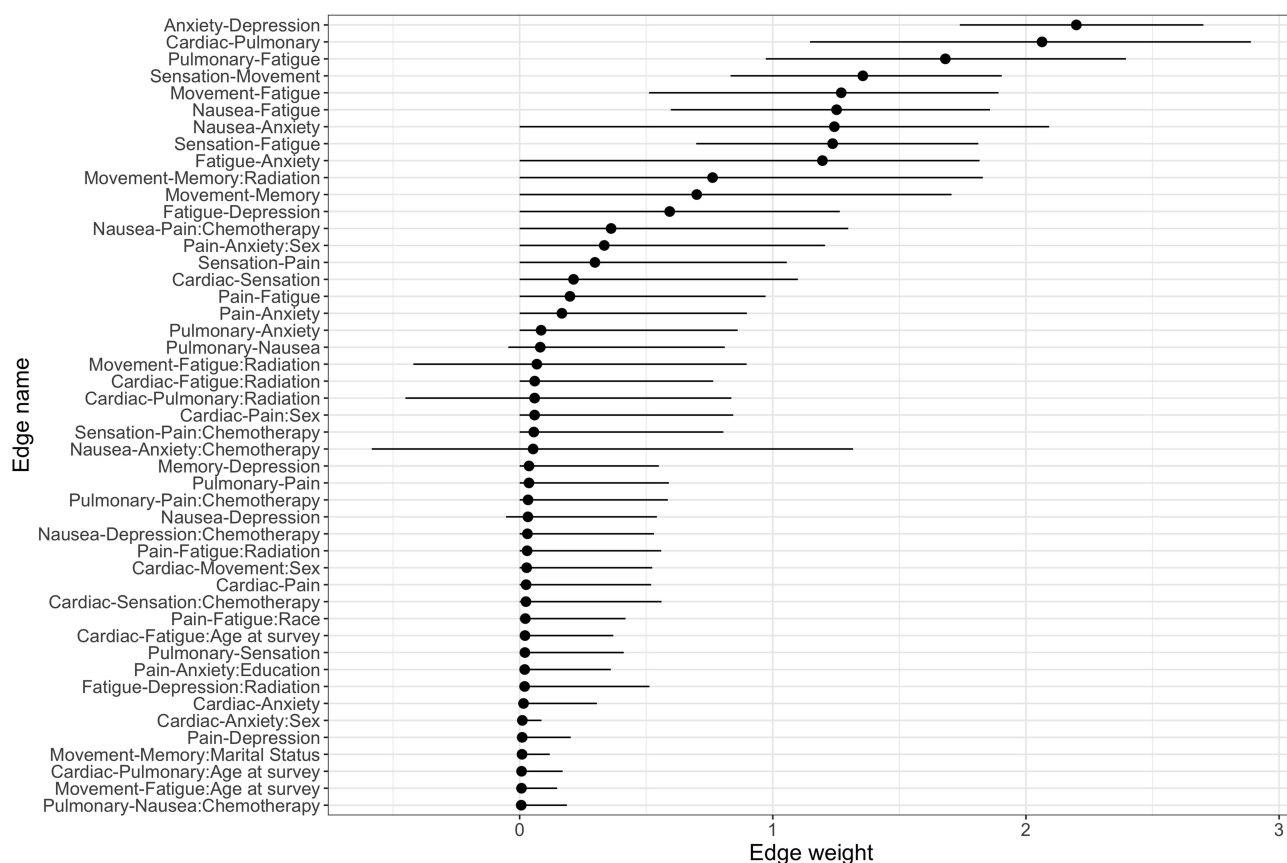
Figure 4 The identified pairwise associations between symptoms of the personal symptom network estimated based on the adult survivors of childhood cancer from SJLIFE. The edge name "Y_1-Y_2" represents the identified association between symptoms Y_1 and Y_2 independent of individual characteristics, and the edge name **"Y_1-Y_2:X"** represents the identified association between symptoms **Y_1** and **Y_2** influenced by individual characteristic **X**. The black dots represent the average point estimates, and the lines represent the corresponding 95% CIs.

a stronger association. Although the 95% CI includes zero, the association is still considered significant because Lasso was used for variable selection in the personal network estimation.[29] Chemotherapy influenced the association between nausea and pain (0.36 95% CI 0–1.30), with survivors who ever received chemotherapy showing a stronger association. Sex influenced associations like pain with anxiety (0.33 95% CI 0–1.21) and cardiac symptoms with pain (0.06 95% CI 0–0.84), with female survivors showing a stronger association compared to male survivors. Age in the survey affected associations like cardiac symptoms with fatigue (0.02 95% CI 0–0.37), with older age indicating stronger associations. Other covariate-related associations were identified, including pain and fatigue affected by race/ethnicity (0.02 95% CI 0–0.42) and pain and anxiety affected by education (0.02 95% CI 0–0.36). Racial/ethnic minority and education above college/post-graduate level showed stronger associations. Marital status influenced the association between movement and memory problems (0.01 95% CI 0–0.12), with survivors who were never-married or lived as married indicating a stronger association. Personal income did not significantly influence the personal symptom network structure.

Symptom networks for individual cancer survivors were constructed based on the identified associations between symptom domains and personal characteristics. Figure 5 depicts the networks for two representative survivors: the first survivor (labeled as "a") characterized as non-Hispanic White, male, 30.9 years of age (overall mean age) at survey with below college/post-graduate education, ever-married or lived as married, and never received chemotherapy and/or radiation, and a second survivor (labeled as "b") characterized as a racial/ethnic minority, female, 39.5 years of age (one SD above the mean) at survey with above college/post-graduate education, never-married or lived as married, and received chemotherapy and/or radiation. Figure 5 displays the individualized symptom profile and suggests that survivor "a", who had less vulnerable sociodemographic and clinical factors, experienced less complex interconnected symptoms compared to survivor "b", who had more vulnerable sociodemographic and clinical factors. For survivor "b", older age had stronger associations between cardiac symptoms and fatigue compared to younger
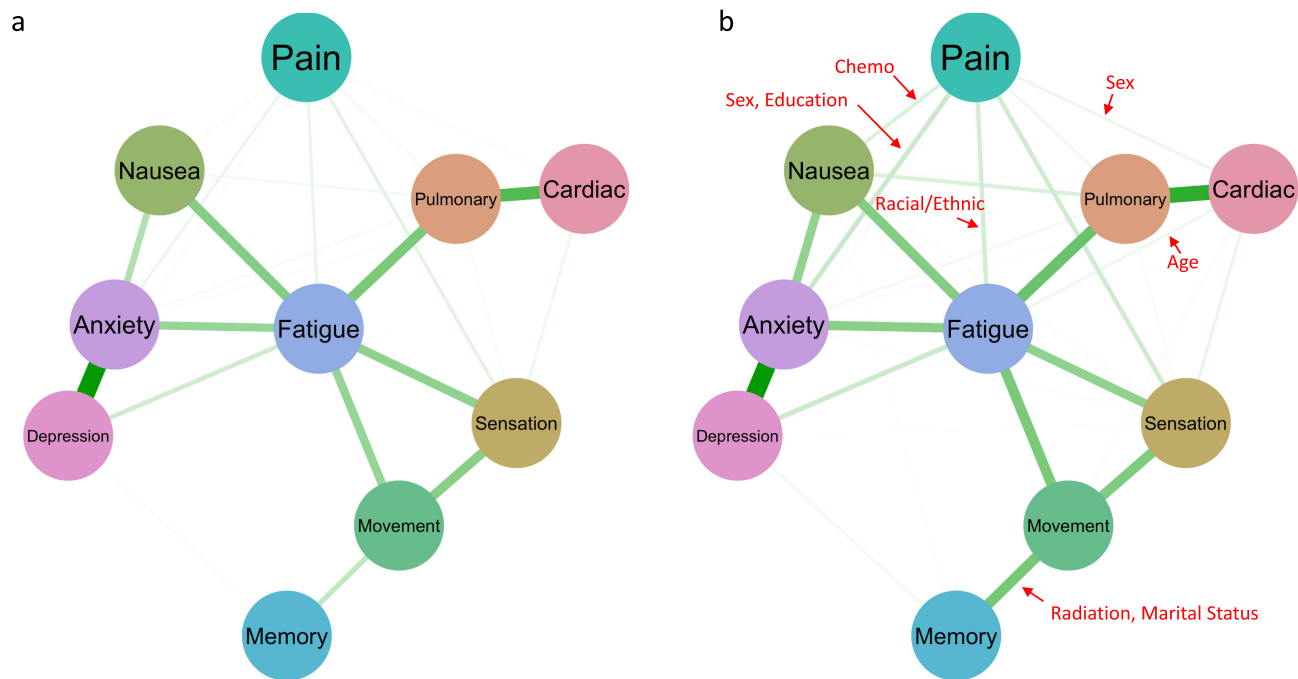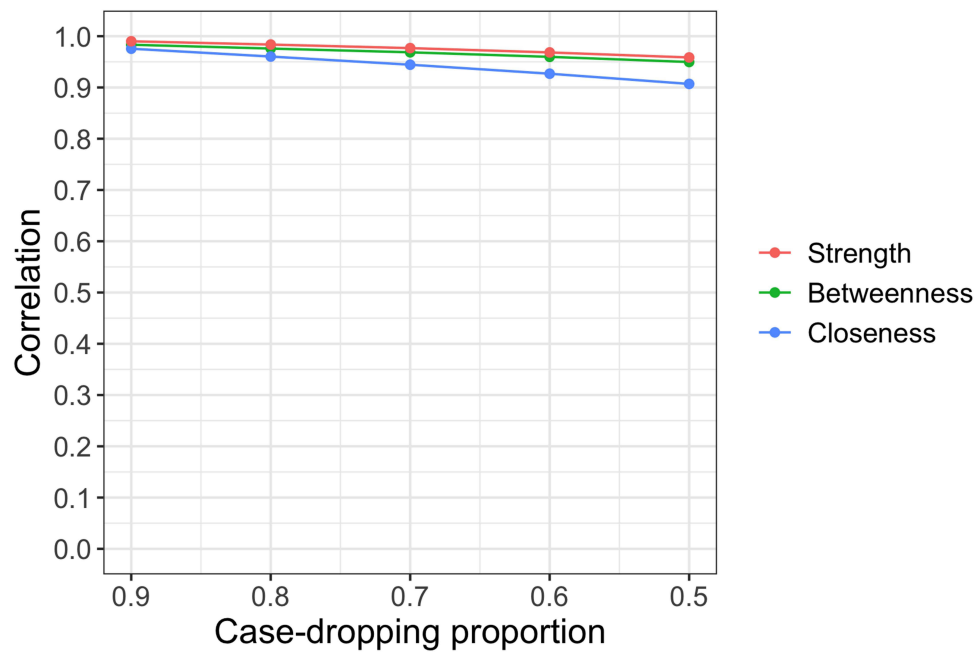
Figure 5 The estimated personal symptom network for two individual cancer survivors: (**a**) a White non-Hispanic male, aged 30.9 years (overall mean age) at the survey with below college/post-graduate education, ever married or lived as married, and never received chemotherapy and/or radiation, and (**b**) a racial/ethnic minority female, aged 39.5 years (one SD above the mean) at the survey with above college/post-graduate education, never married or lived as married, and received chemotherapy and radiation. Edges pointed to by arrows with covariates indicate the influence of covariates on the pairwise associations between symptoms.

age. The female sex exhibited stronger associations between cardiac symptoms and pain, as well as between pain and anxiety compared to the male sex. Racial/ethnic minorities presented greater associations between pain and fatigue, compared to non-Hispanic White. Survivors with above a college education had inflated associations between pain and anxiety compared to below college/post-graduate education. The status of never-married or lived as married had a stronger association between movement and memory problems compared to ever-married or lived as married. Additionally, ever receiving chemotherapy and/or radiation had stronger associations between pain and nausea, and between movement and memory problems, respectively.

Centralities, including strength, betweenness, and closeness, were calculated for nodes in the personal symptom networks of the two representative survivors (Figure 6). The female survivor exhibited higher strength, betweenness, and closeness for all symptom domains, except for the lower betweenness for fatigue. Case-dropping bootstrap was used to assess centrality stability, and Figure 7 illustrates high correlations of centralities even with a 50% random data drop, indicating stable centralities in the estimated personal symptom networks. Finally, we assessed significant differences in associations and
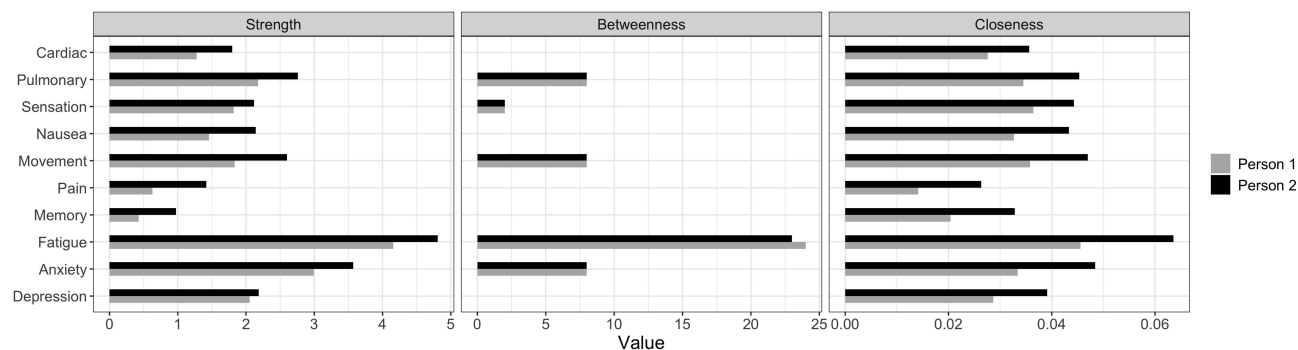
Figure 6 Values of centrality indices (strength, betweenness, and closeness) in the representative personal symptom network for person 1: a White non-Hispanic male, aged 30.9 years at the survey with below college/post-graduate education, ever married or lived as married, and never received chemotherapy and/or radiation, and person 2: a racial/ethnic minority female, aged 39.5 years at survey with above college/post-graduate education, never married or lived as married, and received chemotherapy and radiation.

Figure 7 Correlations of centrality metrics in case-dropping bootstrap testing for the personal symptom network estimated based on the adult survivors of childhood cancer from SJLIFE.

centralities using tests based on empirical distributions derived from bootstrap ([Supplementary Figures 1](#) and [2](#)). Results indicate significant differences between specific associations (eg, anxiety-depression vs cardiac symptoms-pulmonary symptoms) and centralities (eg, strength of pain vs anxiety).

Discussion

In this study, we estimated the influence of individual characteristics on associations among multiple symptoms for adult survivors of childhood cancer and evaluated the performance of the Ising model with covariates to construct personal networks of binary symptom data. Simulation experiments confirmed the model's accuracy and stability in estimating personal symptom networks. Application of the Ising model with covariates to the empirical symptom data revealed age, sex, race/ethnicity, education, marital status, and treatments (any chemo and radiation therapy) as significant factors shaping symptom associations among childhood cancer survivors. Notably, when incorporating covariates into the Ising model, we excluded time from cancer diagnosis for two main reasons. Firstly, this variable was highly correlated with age at the survey. Considering model parsimony, we only included age at the survey and opted not to include both time from cancer diagnosis and age at diagnosis as covariates in the model. Secondly, compared to time from cancer diagnosis, treatment variables indeed played a more important role in influencing the structure of symptom networks for adult survivors of childhood cancer.[2] Therefore, we prioritized the inclusion of treatment (any chemo and radiation therapy) as a covariate in our network estimation model.

Leveraging the constructed personal symptom network, future approaches to symptom management may adopt a dual focus. This entails targeting interventions towards either the individual characteristics or the central symptoms identified within the network. For example, given that sex was identified to play an important role in influencing associations between pain and anxiety, and between cardiac symptoms and pain, future research may consider gender difference in symptom perception or experience into the interventional design for effectively addressing these symptom experiences. Similarly, attention should be directed towards older adult survivors of childhood cancer when delivering interventions for managing fatigue and cardiac symptoms. One interesting finding of our study was the significant impact of education on the association between pain and anxiety. This underscores the importance of providing lay-person/low-literacy educational materials for childhood cancer survivors to better manage the experiences of pain and anxiety.

As mentioned in the "Evaluation of network accuracy and stability" section in the [Supplementary Materials](#), data splitting inference was employed in this work to assess the accuracy and stability of the estimated personal symptom networks. This

approach eliminates bias in coefficient estimation caused by the shrinkage effect of Lasso and solves the problem of divergent model selection that arises from performing Lasso on multiple bootstrapped datasets. Simulation results illustrated that data splitting inference can generate valid empirical distributions of the estimated coefficients in the network and control type I error rates of edge and centrality testing at the desired level. Previous research[26] has demonstrated that eBIC generates consistent model selection results and performs best with $\eta = 0.25$ for the Ising model. It was shown in our simulation that larger η values resulted in more conservative edge identification. Based on the simulation results, we also suggest researchers set $\eta = 0.25$ in personal symptom network modeling.

Major limitations of this work include the inclusion of all binary symptom data and the use of a cross-sectional design to estimate personal symptom networks. Our approach merely captures the presence of symptom attribute rather than the severity or interference attribute of symptoms that are often in a continuous or ordinal in nature. Additionally, the symptom burden of individuals with cancer may change over time, given the progression of late effects post-cancer therapy or the aging process. One future direction of this research is to extend the current model to estimate personal symptom networks from continuous or ordinal symptom data and to build temporal networks to study the dynamic changes of network structures over time, which may deepen our understanding of the symptom network dynamics, facilitate identification of symptom etiology, and design of personalized interventions to improve symptom management.

Conclusion

By applying simulation methods to evaluate the Ising model with covariates for estimating personal symptom networks and employing the real-world data from childhood cancer survivors to assess these networks across 10 symptom domains, we reveal that individual differences have a substantial impact on the manifestation of a variety of symptoms observed within the networks. Our findings shed light on the complex relationships between various symptoms that individual cancer survivors face, offering insights with significant implications for clinical practice. By identifying central symptoms within these networks, our research offers valuable insights for further identifying the biological mechanisms and targeted interventions to improve symptom management among childhood cancer survivors.

Abbreviation

CI, confidence interval; eBIC, extended Bayesian information criterion; FPR, false positive rate; MCC, Matthew's correlation coefficient; SJLIFE, St. Jude Lifetime Cohort Study; TPR, true positive rate; SD, standard deviation.

Data Sharing Statement

The deidentified data will be publicly available on Zenodo.org.

Funding

Disclosure

I-Chan Huang and Deokumar Srivastava are co-senior authors for this study. The authors declare no conflicts of interest in this work.

References

1. Hong HC, Min A, Kim YM. A systematic review and pooled prevalence of symptoms among childhood and adolescent and young adult cancer survivors. *J Clin Nurs*. 2023;32(9–10):1768–1794. doi:10.1111/jocn.16201
2. Shin H, Dudley WN, Bhakta N, et al. Associations of symptom clusters and health outcomes in adult survivors of childhood cancer: a report from the St Jude Lifetime Cohort Study. *J Clin Oncol*. 2023;41(3):497–507. doi:10.1200/JCO.22.00361
3. Huang I-C, Brinkman TM, Kenzik K, et al. Association between the prevalence of symptoms and health-related quality of life in adult survivors of childhood cancer: a report from the St Jude Lifetime Cohort study. *J Clin Oncol*. 2013;31(33):4242. doi:10.1200/JCO.2012.47.8867
4. McNally RJ. Can network analysis transform psychopathology? *Behav Res Therap*. 2016;86:95–104. doi:10.1016/j.brat.2016.06.006
5. Barzel B, Barabasi AL. Universality in network dynamics. *Nat Phys*. 2013;9:673–681. doi:10.1038/nphys2741
6. Liu YY, Slotine JJ, Barabasi AL. Controllability of complex networks. *Nature*. 2011;473(7346):167–173. doi:10.1038/nature10011

7. Strogatz SH. Exploring complex networks. *nature*. 2001;410(6825):268–276. doi:10.1038/35065725

8. Dorogovtsev SN, Mendes JF. *Evolution of Networks: From Biological Nets to the Internet and WWW*. Oxford university press; 2003.

9. Albert R, Barabási A-L. Statistical mechanics of complex networks. *Rev Mod Phys*. 2002;74(1):47. doi:10.1103/RevModPhys.74.47

10. Papachristou N, Barnaghi P, Cooper B, et al. Network analysis of the multidimensional symptom experience of oncology. *Sci Rep*. 2019;9(1):2258. doi:10.1038/s41598-018-36973-1

11. Cramer AO, Waldorp LJ, van der Maas HLJ, et al. Comorbidity: a network perspective. *Behav Brain Sci*. 2010;33(2–3):137–150. doi:10.1017/S0140525X09991567

12. Borsboom D, Cramer AO. Network analysis: an integrative approach to the structure of psychopathology. *Annual Rev Clin Psych*. 2013;9:91–121. doi:10.1146/annurev-clinpsy-050212-185608

13. Borsboom D. Psychometric perspectives on diagnostic systems. *J Clin Psych*. 2008;64(9):1089–1108. doi:10.1002/jclp.20503

14. Borgatti SP, Mehra A, Brass DJ, et al. Network analysis in the social sciences. *science*. 2009;323(5916):892–895. doi:10.1126/science.1165821

15. Van Der Maas HL, Dolan CV, Grasman RPPP, et al. A dynamical model of general intelligence: the positive manifold of intelligence by mutualism. *Psychol Rev*. 2006;113(4):842. doi:10.1037/0033-295X.113.4.842

16. Shim EJ, Ha H, Suh Y-S, et al. Network analyses of associations between cancer-related physical and psychological symptoms and quality of life in gastric cancer patients. *Psycho-Oncology*. 2021;30(6):946–953. doi:10.1002/pon.5681

17. Kalantari E, Kouchaki S, Miaskowski C, et al. Network analysis to identify symptoms clusters and temporal interconnections in oncology patients. *Sci Rep*. 2022;12(1):17052. doi:10.1038/s41598-022-21140-4

18. Oeffinger KC, Mertens AC, Sklar CA, et al. Chronic health conditions in adult survivors of childhood cancer. *N Engl J Med*. 2006;355(15):1572–1582. doi:10.1056/NEJMsa060185

19. Cheng J, Levina E, Wang P, et al. A sparse Ising model with covariates. *Biometrics*. 2014;70(4):943–953. doi:10.1111/biom.12202

20. Haslbeck J. Estimating group differences in network models using moderation analysis. *PsyArXiv*. 2020.

21. Haslbeck JM, Borsboom D, Waldorp LJ. Moderated network models. *Multivariate Behavioral Research*. 2021;56(2):256–287. doi:10.1080/00273171.2019.1677207

22. Qian M, Murphy SA. Performance guarantees for individualized treatment rules. *Ann Stat*. 2011;39(2):1180. doi:10.1214/10-AOS864

23. Molenaar PC. A manifesto on psychology as idiographic science: bringing the person back into scientific psychology, this time forever. *Measurement*. 2004;2(4):201–218.

24. Epskamp S. Psychometric network models from time-series and panel data. *Psychometrika*. 2020;85(1):206–231. doi:10.1007/s11336-020-09697-3

25. Epskamp S, Hoekstra RH, Burger J, et al. Longitudinal Design choices: Relating data to analysis. In: *Network Psychometrics with R*. Routledge; 2022:157–168.

26. Foygel R, Drton M. Extended bayesian information criteria for gaussian graphical models. *Adv Neural Inform Processi Syst*. 2010;23:23.

27. Barber RF, Drton M, *High-dimensional Ising model selection with Bayesian information criteria*. 2015.

28. Van Borkulo CD, Borsboom D, Epskamp S, et al. A new method for constructing networks from binary data. *Sci Rep*. 2014;4(1):5918. doi:10.1038/srep05918

29. Epskamp S, Borsboom D, Fried EI. Estimating psychological networks and their accuracy: a tutorial paper. *Behav Res Meth*. 2018;50:195–212. doi:10.3758/s13428-017-0862-1

30. Koller D, Friedman N. *Probabilistic Graphical Models: Principles and Techniques*. MIT press; 2009.

31. Loh P-L, Wainwright MJ. Structure estimation for discrete graphical models: Generalized covariance matrices and their inverses. *Adv Neural Inform Process Syst*. 2012;25.

32. Ravikumar P, Wainwright MJ, Lafferty JD, *High-dimensional Ising model selection using ℓ 1-regularized logistic regression*. 2010.

33. Marsman M, Huth K, Waldorp LJ, et al. Objective Bayesian edge screening and structure selection for ising networks. *psychometrika*. 2022;87(1):47–82. doi:10.1007/s11336-022-09848-8

34. Epskamp S, Fried EI. A tutorial on regularized partial correlation networks. *Psychological Methods*. 2018;23(4):617. doi:10.1037/met0000167

35. Tibshirani R. Regression shrinkage and selection via the lasso. *J Royal Stat Societ Series B*. 1996;58(1):267–288. doi:10.1111/j.2517-6161.1996.tb02080.x

36. Chen J, Chen Z. Extended Bayesian information criteria for model selection with large model spaces. *Biometrika*. 2008;95(3):759–771. doi:10.1093/biomet/asn034

37. Meinshausen N, Bühlmann P, *High-dimensional graphs and variable selection with the lasso*. 2006.

38. Rinaldo A, Wasserman L, G'Sell M, *Bootstrapping and sample splitting for high-dimensional, assumption-lean inference*. 2019.

39. Howell CR, Bjornard KL, Ness KK, et al. Cohort profile: The St. Jude lifetime cohort study (SJLIFE) for paediatric cancer survivors. *Int J Epidemiol*. 2021;50(1):39–49. doi:10.1093/ije/dyaa203

40. Finnemann A, Borsboom D, Epskamp S, et al. The theoretical and statistical Ising model: a practical guide in R. *Psych*. 2021;3(4):593–617. doi:10.3390/psych3040039