Title: Multicenter Study on Symptom Cluster Heterogeneity and Related Gut Microbiota and Metabolite Mechanisms in Childhood Cancer Survivors

NCT Number: NCT07450872

Date: February 4, 2026

Research Protocol

Background

With advances in diagnostic and therapeutic technologies, an increasing number of pediatric cancer patients can achieve long-term survival following standardized treatment. The 5-year net survival rate for common childhood cancers has reached 80–95% in high-income regions such as Europe, North America, and East Asia. In China, the long-term survival rate of childhood cancer is slightly lower than that in developed countries, with an overall 5-year relative survival rate of 71.9%. The improved survival rate has led to a growing population of **childhood cancer survivors (CCS)**.

Despite improved long-term survival, the impact of cancer and its treatment on CCS cannot be overlooked. Treatments such as chemotherapy, radiotherapy, surgery, and targeted immunotherapy prolong survival but also cause varying degrees of physiological damage. Studies have shown that CCS often experience 10–16 concurrent physical and psychological symptoms, among which fatigue, pain, insomnia, depression, and nausea are the most common. These symptoms typically occur in clusters, where interactions among symptoms may produce synergistic and amplifying effects, leading to greater negative outcomes. This further exacerbates symptom burden, significantly affects treatment outcomes and health-related quality of life, and may even shorten survival.

Symptom management has become a critical component of CCS care. Understanding the interaction mechanisms among multiple symptoms is of great clinical significance for healthcare providers.

Symptom cluster identification is a classical dimensionality reduction approach in clinical practice aimed at simplifying complex symptom interactions. According to Kim et al., a symptom cluster consists of two or more concurrent and interrelated symptoms that may share common or distinct etiologies. Symptom clusters are relatively independent, while symptoms within a cluster may reinforce or accelerate each other or promote the development of other symptoms. Compared with single symptoms, the combined effects of symptom interactions lead to more severe negative outcomes.

Existing studies indicate that although symptom clusters in CCS show heterogeneity due to differences in cancer types, treatment stages, assessment tools, and analytical methods, core symptom cluster types demonstrate cross-study stability. For example, Fang et al. constructed a symptom network including 31 symptoms among children undergoing chemotherapy and identified five clusters: mucocutaneous symptoms, physical symptoms, self-image disturbance, gastrointestinal symptoms, and

psychological symptoms. These findings are consistent with studies by Yeh et al. and Hockenberry et al.

In recent years, research on childhood cancer symptoms in China has shifted from cross-sectional to longitudinal designs, aiming to explore dynamic changes in symptom clusters across disease stages. For instance, longitudinal tracking of children with acute lymphoblastic leukemia showed that although the number of symptom clusters varied across treatment stages (3–6 clusters), physical, gastrointestinal, and psychological symptom clusters consistently existed, indicating temporal stability of core clusters.

Based on existing evidence, symptom clusters in pediatric cancer can be broadly categorized into three types:**Disease- and treatment-related symptom clusters** (e.g., malnutrition, cardiopulmonary toxicity, motor dysfunction, cognitive impairment, mucosal damage); **Physical symptom clusters** (e.g., pain, fatigue, sleep disturbance, nausea, vomiting, anorexia); **Psychological symptom clusters** (e.g., anxiety, depression, irritability, worry).

These clusters are repeatedly identified across different diseases and treatment stages, suggesting long-term stability and potential biologically driven mechanisms rather than short-term treatment effects.

However, current research has several limitations:

Stability verification relies mainly on short-term repeated measurements (≤3 months) or cross-sectional comparisons;

Heterogeneity analyses are limited to single cancer types or treatment stages;

Definitions of core symptoms lack standardized criteria and quantitative cross-time stability measures.

Therefore, this study will employ repeated measurements at baseline, 3 months, and 6 months, combined with exploratory factor analysis and latent growth mixture modeling, to systematically validate symptom cluster stability across cancer types and treatment stages.

Potential mechanisms underlying symptom cluster stability may involve:

Psychosocial factors (e.g., personal traits, psychological status, social environment);**Pathophysiological processes** (e.g., chemotherapy-induced mucosal injury, neuroinflammation).

In addition, growing evidence from epidemiology and microbiome research suggests that gut microbiota and their metabolites play important roles in mediating disease

and behavioral outcomes. Gut microbiota influence host metabolism, immune regulation, and inflammation through pathways such as the gut–brain axis, gut–lung axis, gut–bone axis, and others. These findings provide new insights into the biological mechanisms underlying symptom clusters in pediatric cancer.

Objectives

To describe the prevalence and severity of common symptoms in CCS, identify symptom cluster types, and validate their longitudinal stability using baseline and follow-up data.

To explore heterogeneity in symptom cluster trajectories; identify demographic, physiological, psychological, and social factors influencing trajectory patterns; quantify their relative importance; and examine their interactions and predictive effects on quality of life.

To compare gut microbiota composition and diversity across different symptom cluster trajectory subgroups, identify characteristic taxa and metabolites associated with high symptom burden, and explore underlying microbiome-related mechanisms.

Methods

Phase I: Identification and Validation of Symptom Clusters

A longitudinal design will be used. Field surveys will assess symptom prevalence and severity. Exploratory factor analysis will identify symptom clusters and core symptoms based on baseline data. Cross-sample and longitudinal validation will be conducted to assess stability over time.

Phase II: Trajectories, Influencing Factors, and Quality of Life

Latent growth mixture modeling will be used to identify symptom cluster trajectories. Logistic regression and machine learning algorithms will determine influencing factors and their relative importance. Network analysis will examine temporal interactions. Negative binomial and zero-inflated Poisson models will evaluate the predictive effects on quality of life.

Phase III: Gut Microbiota and Metabolomics Mechanisms

A nested case–control study will be conducted. Biological samples will be collected for:16S rRNA sequencing (microbiome analysis); Liquid chromatography–mass spectrometry (metabolomics)

Differences in microbiota composition, diversity, and metabolites between trajectory subgroups will be analyzed. Machine learning methods will identify key microbial taxa and metabolites.

Sample Size

Sample size is determined based on statistical requirements:

Phase I: ≥200 participants (for stable factor analysis results)

Phase II: Based on latent growth models, regression, and network analysis requirements

Phase III: Derived from Phases I and II

Considering feasibility and potential attrition, the total sample size is estimated at **approximately 400 participants**.

Study Procedures

Phase I

Baseline survey conducted during hospitalization, followed by 3- and 6-month follow-ups. Data collected include demographics, clinical information, and symptom assessments.

Phase II

Participants completing all follow-ups will be included. Only non-invasive variables (e.g., demographics, clinical data, lifestyle factors) will be analyzed.

Phase III

Biological samples will be collected under fasting conditions. Case and control groups will be defined based on symptom trajectory patterns. Multi-omics and statistical analyses will be conducted.

Outcome Measures

Primary Outcome Measure

The primary outcome is the severity and longitudinal trajectory of symptom clusters among childhood cancer survivors. Symptoms will be assessed using the Memorial Symptom Assessment Scale 10-18 (MSAS 10-18) at baseline, 3 months, and 6 months. Symptom occurrence, frequency, severity, and distress will be used to identify symptom clusters and core symptoms. Exploratory factor analysis, latent growth mixture modeling, and network analysis will be used to examine symptom cluster structure, trajectory heterogeneity, and stability over time.

Secondary Outcome Measures

Secondary outcomes include health-related quality of life, psychosocial factors, lifestyle factors, gut microbiota composition, and fecal metabolite profiles. Quality of life will be assessed using the Pediatric Quality of Life Inventory Cancer Module and Generic Core Scales. Psychosocial variables will include anxiety, depression, perceived stress, self-efficacy, post-traumatic growth, coping style, parent-child closeness, parenting style, and family functioning. Gut microbiota will be assessed using 16S rRNA sequencing, and fecal metabolites will be assessed using liquid chromatography-mass spectrometry.

Eligibility Criteria

Eligible participants are childhood cancer survivors aged 8-18 years with a clinically and pathologically confirmed diagnosis of cancer, including leukemia, lymphoma, central nervous system tumors, or other common solid tumors. Participants must have been diagnosed for more than 6 months and be in the maintenance/recovery phase or have completed treatment for at least 3 months. Participants must have no cognitive impairment and must be able to understand, communicate, and complete questionnaires. Written informed consent from legal guardians and assent from children will be obtained before enrollment.

Participants will be excluded if they have severe treatment-related sequelae that may interfere with symptom assessment, poor compliance or inability to complete questionnaires independently, concurrent participation in other interventional clinical trials, or planned major surgery or changes in chemotherapy/radiotherapy during the study period.

Sample Size

The planned sample size is 400 participants. This estimate is based on the requirements of exploratory factor analysis, latent growth mixture modeling, regression analysis, and network analysis, while allowing for an estimated 10% attrition rate.

Statistical Analysis

Descriptive statistics will be used to summarize demographic and clinical characteristics. Continuous variables will be presented as means and standard deviations or medians and interquartile ranges, as appropriate. Categorical variables will be presented as frequencies and percentages.

Exploratory factor analysis will be used to identify symptom clusters based on baseline symptom severity scores. The Kaiser-Meyer-Olkin test and Bartlett’s test of sphericity will be used to assess the suitability of the data for factor analysis. Factor extraction will be guided by eigenvalues, scree plots, factor loadings, and clinical interpretability.

Longitudinal stability of symptom clusters will be examined using repeated assessments at baseline, 3 months, and 6 months. Latent growth mixture modeling will be used to identify heterogeneous symptom cluster trajectory subgroups. Model selection will be based on information criteria, entropy, likelihood ratio tests, class size, and clinical interpretability.

Multivariable logistic regression and machine learning methods will be used to identify factors associated with symptom cluster trajectory membership. Candidate predictors will include demographic characteristics, clinical factors, lifestyle variables, psychological factors, and social environment variables. Network analysis and cross-lagged network analysis will be used to examine relationships among symptoms and influencing factors over time.

Quality of life outcomes will be analyzed using regression models adjusted for relevant demographic, clinical, and psychosocial covariates. Appropriate models will be selected according to the distribution of outcome variables.

For the nested case-control analysis, participants with high symptom burden trajectories will be selected as cases, and participants with low and stable symptom trajectories will be selected as controls. Gut microbiota alpha diversity, beta diversity, relative abundance, and taxonomic composition will be compared between groups. Differential metabolites will be identified using metabolomics analysis. Random forest models and multivariable regression models will be used to identify key microbial taxa and metabolites independently associated with high symptom burden after adjustment for relevant covariates.

Missing data will be assessed for extent and pattern. If appropriate, multiple imputation or full-information maximum likelihood methods will be used. Sensitivity analyses will be conducted to examine the robustness of the findings. Statistical significance will be set at a two-sided P value of less than 0.05. Adjustment for multiple comparisons will be applied where appropriate, particularly for microbiome and metabolomics analyses.