BMJ Open

Risk factors and prediction model of breast cancer-related lymphoedema in a Chinese cancer centre: a prospective cohort study protocol

Aomei Shen ,[1,2] Jingming Ye,[3] Hongmei Zhao,[4] Wanmin Qiang,[1] Hongmeng Zhao,[5] Yubei Huang,[6] Yujie Zhou,[7] Yue Wang,[3] Xin Li,[1,8] Zhongning Zhang,[1,8] Jingru Bian,[1,8] Liyuan Zhang,[1,8] Peipei Wu,[9] Ying Wang,[1] Qian Lu [2]

To cite: Shen A, Ye J, Zhao H, *et al*. Risk factors and prediction model of breast cancer-related lymphoedema in a Chinese cancer centre: a prospective cohort study protocol. *BMJ Open* 2024;**14**:e089769. doi:10.1136/bmjopen-2024-089769

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (https://doi.org/10.1136/bmjopen-2024-089769).

Received 10 June 2024
Accepted 26 November 2024

For numbered affiliations see end of article.

Correspondence to
Professor Qian Lu;
luqian@bjmu.edu.cn

ABSTRACT

Introduction Lymphoedema is a distressing and long-term complication for breast cancer survivors. However, the reported incidence of lymphoedema varies, and its risk factors remain underexplored. Currently, a well-established risk prediction model is still lacking. This study aims to describe the rationale, objectives, protocol and baseline characteristics of a prospective cohort study focused on examining the incidence and risk factors of breast cancer-related lymphoedema (BCRL), as well as developing a risk prediction model.

Methods and analysis This study is an ongoing single-centre prospective observational cohort study recruiting 1967 patients with breast cancer scheduled for surgery treatment in northern China between 15 February 2022 and 21 June 2023. Assessments will be conducted presurgery and at 1, 3, 6, 12, 18, 24, 30 and 36 months postsurgery. Bilateral limb circumferences will be measured by patients at home or by researchers at the outpatient clinics during follow-up visits. The diagnosis of lymphoedema is based on a relative limb volume increase of ≥10% from the preoperative assessment. Self-reported symptoms will be assessed to assist in diagnosis. Potential risk factors are classified into innate personal traits, behavioural lifestyle, interpersonal networks, socioeconomic status and macroenvironmental factors, based on health ecology model. Data collection, storage and management were conducted using the online 'H6WORLD' data management platform. Survival analysis using the Kaplan-Meier estimate will determine the incidence of BCRL. Risk factors of BCRL will be analysed using log-rank test and COX-LASSO regression. Traditional COX regression analysis and seven common survival analysis machine learning algorithms (COX, CARST, RSF, GBSM, XGBS, SSVM and SANN) will be employed for model construction and validation.

Ethics and dissemination The study protocol was approved by the Biomedical Ethics Committee of Peking University (IRB00001052-21124) and the Research Ethics Committee of Tianjin Medical University Cancer Institute and Hospital (bc2023013). The results of this study will be published in peer-reviewed journals and will be presented at several research conferences.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ This study followed a rigorous process to identify potential factors associated with breast cancer-related lymphoedema (BCRL), combining evidence from systematic reviews, clinical experiences and expert opinions.
⇒ The health ecology model is introduced to help classify multilayer factors of BCRL, including innate personal traits, behavioural lifestyle, interpersonal networks, socioeconomic status and macroenvironmental factors.
⇒ To our knowledge, this study is the first to use patients' self-measurement for outcome assessment in a prospective cohort of BCRL.
⇒ One limitation of this study protocol is that all patients are recruited from a single site, potentially restricting the external validity.
⇒ Potential biomarkers or genetic indicators associated with the development of BCRL were not included considering clinical applicability.

Trial registration number ChiCTR2200057083.

INTRODUCTION

Breast cancer-related lymphoedema (BCRL) is a major concern for breast cancer (BC) survivors. Studies report varying incidence of BCRL, ranging from 4.5% to 58.8%, with a pooled incidence of 21.9% (95% CI 19.8% to 24.0%) across 84 cohort studies.[1] Treatments such as axillary lymph node dissection (ALND), axillary radiotherapy and chemotherapy are identified as the main causes of BCRL. Although advancements in less invasive procedures like sentinel lymph node biopsy (SLNB) and more precise radiotherapy have been associated with a reduced incidence of BCRL, the risk of developing this condition persists.[2 3] Currently, lymphoedema remains incurable. Its onset is a distressing and

enduring event that significantly affects patients' lives.[4] The impact of BCRL is multifaceted, affecting individuals physically, psychosocially and economically.[5 6] Patients with BCRL experience a variety of distressing symptoms, including pain, swelling and numbness, etc.[7] Additionally, they often suffer from altered body image, social isolation, reduced work capacity and the emotional distress of managing long-term lymphoedema treatment, etc, all of which considerably diminish their quality of life.[8 9]

The characteristics and adverse effects of BCRL underscore the importance of its early prediction and prevention. Significant research has been dedicated to identifying risk factors for BCRL and predicting the risk of developing lymphoedema. Established risk factors include treatment-related factors such as ALND, more lymph nodes dissected, radiotherapy, taxane-based chemotherapy and postoperative complications like infections and seroma; personal factors, such as higher body mass index (BMI) and older age, along with disease-related factors like advanced tumour stages.[1 10] A systematic review of current studies on BCRL risk factors highlighted the need to focus on modifiable psychosocial-behavioural factors from a Health Ecology perspective.[1 10 11] Research has shown associations between the development of BCRL and psychosocial-behavioural factors, including insufficient exercise, residence in metropolitan areas and poor financial status.[2 12 13]

Prediction models hold great potential to improve healthcare decision-making by accurately identifying patients who require comprehensive preventive interventions, such as regular monitoring, exercise, weight management and the proactive use of compression sleeves, etc. These models also play a vital role in educating patients with BC and their families about the risks of lymphoedema, enhancing awareness and fostering compliance with preventive measures.[14] Despite significant research efforts over the past decade to develop risk prediction models for BCRL, their integration into clinical practice has been limited. A systematic review of 17 BCRL prediction models showed that these models were often inadequately reported, prone to significant bias, and their performance was often overly optimistic. Consequently, none of these models currently meet the criteria for recommendation in clinical practice guidelines. This situation highlights a critical research gap and underscores the need for developing robust, well-validated prediction models that can reliably be used in clinical settings to effectively prevent and manage lymphoedema.[15] Hence, there is a need for a well-designed prospective cohort study to determine the risk factors and develop a risk prediction model of BCRL, from the perspective of health ecology model.

METHODS AND ANALYSIS

Aims

This prospective cohort study aims (1) to determine the incidence of BCRL within the first 3 years postsurgery among ipsilateral BC women; (2) to observe the interlimb arm volume difference change; (3) to identify the risk factors of BCRL from the perspective of health ecology model and (4) to develop and validate a risk prediction model of BCRL.

Development of the protocol

This study protocol was developed based on systematically reviewing the current research on risk factors and risk prediction models of BCRL.[1 10 15] After examining and assessing 17 risk prediction models of BCRL, we found that the current prediction models were at high risk of bias and poorly reported. Moreover, most existing models focus on unmodifiable disease and treatment-related factors, such as radiotherapy, BMI before surgery, the number of lymph nodes dissected and chemotherapy.[15] There is a notable lack of consideration for psychosocial, behavioural and environmental factors, which are crucial for understanding the multifactorial nature of BCRL. For instance, a recent study has identified rurality as a critical risk factor for BCRL.[16] Recognising that health results from the interaction between individual characteristics and contextual factors,[17] this study introduces the health ecology model to help classify and understand multilayer factors of BCRL, including innate personal traits, behavioural lifestyle, interpersonal networks, socioeconomic status and macroenvironmental factors.[1 10] We followed a rigorous process to identify potential factors associated with BCRL, combing evidence from systematic reviews, clinical experiences and experts' opinions.

Design and setting

This is a prospective cohort study conducted at the Breast Cancer Prevention and Treatment Research Center within a national cancer institute and hospital, a National Clinical Research Center for Cancer, which is one of the largest specialised cancer hospitals in the country. The Breast Cancer Prevention and Treatment Research Center features 600 beds and serves an annual outpatient volume of 360 000 visits from across the country, in addition to performing more than 8000 surgeries each year. The recruitment for this study began on 15 February 2022 and continued until 21 June 2023. The follow-up will last until 21 June 2026. The study protocol has been registered on Chinese Clinical Trial Registry (Registration number: ChiCTR2200057083; https://www.chictr.org.cn/bin/project/edit?pid=135245).

Participants and eligibility

This study recruited females with unilateral BC who underwent unilateral radical mastectomy or modified radical mastectomy between February 2022 and June 2023. Inclusion criteria were as follows: (1) patients newly diagnosed with BC via pathology; (2) patients who have undergone radical mastectomy or modified radical mastectomy; (3) female; (4) age ≥18 years; (5) mentally clear, without cognitive impairments or communication disorders and (6) informed consent obtained.

The following patients were excluded: (1) patients with bilateral BC; (2) individuals with malignant tumours in non-breast locations; (3) patients who have experienced tumour recurrence or metastasis; (4) patients previously diagnosed with lymphoedema or other lymphatic system diseases; (5) patients with a history of psychiatric disorders; (6) patients suffering from other severe physical conditions (such as severe hypertension (defined as a systolic blood pressure of ≥180 mm Hg and/or diastolic blood pressure of ≥110 mm Hg) or severe heart disease) that hinder participation in the study; (7) patients who are unable to undergo arm circumference measurements due to congenital or acquired arm abnormalities (eg, deformities, amputations, tumours, scars) and (8) patients and their family members who are unable to use smartphones or WeChat to complete questionnaires.

Sample size and power analysis

Based on the 10-fold events per variable principle,[18 19] this study aims to develop a BCRL risk prediction model with 20 predictive variables, aiming to enrol 1014–1143 patients based on a 21.9% of BCRL incidence and a 10%–20% drop-out rate. We intend to include approximately 1150 subjects for model construction and internal validation. Evidence suggests that to achieve 80% statistical power and an alpha of 0.05 for external validation, at least 100 positive and 100 negative cases are needed.[20] Considering the same BCRL incidence and drop-out rate, 508–572 patients should be enrolled, with about 550 participants targeted for external validation. In total, the sample size of this study is approximately 1700.

Outcomes and measures

The primary outcomes were the diagnosis of BCRL and the time of lymphoedema occurrence. Both objective limb circumference measurement and subjective self-reported symptoms will be used to assess lymphoedema.[21 22]

1. Lymphoedema-related symptom assessment: As suggested by prior research, the patient's self-reported lymphoedema-related symptoms could predict the presence of lymphoedema.[23] Evidence and guideline recommended to incorporate self-reported symptoms into the assessment and diagnosis of lymphoedema.[22] Hence, the patient's self-reported lymphoedema-related symptoms was assessed and regarded as an indicator for further objective assessment. The Chinese-version Breast Cancer and Lymphoedema Symptom Experience Index (BCLE-SEI) is used for symptom assessment.[24] The BCLE-SEI consists of two parts: symptom occurrence and symptom distress. In this study, we only used the part I symptom occurrence, which assesses 24 symptoms including limited (shoulder, elbow, wrist, fingers and arm) movement, arm-swelling, breast-swelling, chest wall-swelling, stiffness, tightness, heaviness, fibrosis (toughness or thickness of skin), firmness, tenderness, hotness, redness, blistering, pain/aching/soreness, numbness, burning, stabbing, tingling, fatigue and weakness. Symptoms were assessed using a 4-point Likert-type scale, with scores ranging from 0 (absent) to 4 (severe). The BCLE-SEI was translated into Chinese and validated for the Chinese population, demonstrating good reliability and validity.[24]
2. Bilateral limb measurement: Limb circumference measurements, which are reported as the most clinically used and accessible method, were used to measure and monitor changes in limb.[25 26] Measurement could be performed by research nurses, patients themselves or their trained caregivers.[27] We adhered to a well-established protocol for circumference measurement of the bilateral arms at 10 cm intervals.[21] Using a retractable non-stretch soft tape measure, which facilitates patient self-measurement, measurements began at the midpoint of the ulna styloid, designated as the '0 cm' mark and continued at 10 cm intervals up to 40 cm proximal to the ulna styloid (online supplemental figure 1). Before taking measurements, have the patient extend the arm to be measured. Place the zero point of a paper, adhesive strip measuring tape at the midpoint of the ulna styloid and secure it to the arm. This method enhances the precision of marking intervals and the ease of self-measurement. The arm volume will be calculated using the formula for the volume of a truncated cone: $V=h(C_1^2+C_{1C2}+C_2^2)/12\varpi$, where h represents the height of the cone, which is 10 cm in this study, and C_1 and C_2 symbolise the circumferences of the upper and lower bases of the cone, respectively. The volumes from each section of the arm will then be summed to obtain the total calculated volume.[28]
3. Diagnosis of lymphoedema: The circumferential measure will be converted into volume using the relative volume change (RVC) formula: $RVC\,(\%)=[A_2U_1/A_1U_2-1]\times 100\%$, where A_1 and A_2, U_1 and U_2 are arm volumes of the affected and unaffected arms at preoperative baseline and follow-up, respectively.[29] RVC increase ≥10% from preoperative baseline will be used to diagnose clinical BCRL.[22] RVC increase between 5% and 10% will be regarded as subclinical lymphoedema.[22]

Potential risk factor variables and measures

We conducted a systematic review of risk prediction models for BCRL,[15] an umbrella review of risk factors for BCRL,[10 11] and a meta-analysis of risk factors for BCRL[1] to identify potential risk factors for the development of BCRL. Then, we categorised potential risk factors into five groups based on health ecology model (see figure 1): innate personal traits, behavioural lifestyle, interpersonal networks, socioeconomic status and macroenvironmental factors. Initial sets of potential risk factor variables were developed through discussions within our research group.[1 10 11] Subsequently, the collection of risk factor variables was revised and finalised, considering their importance, relevance and clinical feasibility, through an online expert consultation meeting. 17 experts (average years of experience in this field: 11.53±6.19), specialising in BCRL management and research, participated in this

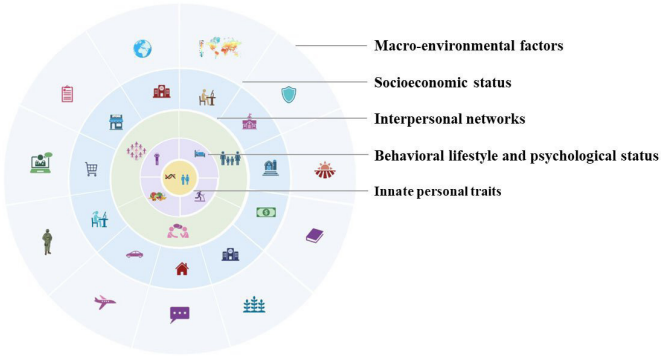

Figure 1 Health ecology model.

consultation. Consensus was reached on a set of risk factor variables (see online supplemental table 1). Actually, we incorporated additional variables based on the consensus to ensure a thorough exploration. The variables and instruments used are as follows.

1. Sociodemographic information: This included age, marital status, education level, dominant hand, employment status, family monthly income, medical insurance, residential condition and accessibility to medical services (timeliness of seeking medical care and overall ease of accessing medical care), exercise habits, etc.
2. Disease and treatment-related information: This included family history of BC, family history of lymphoedema, tumour size, tumour location, tumour clinical stage, TNM stage, type of breast surgery, type of axillary surgery, the number of lymph nodes dissected, the number of positive lymph nodes, adjuvant treatment (neoadjuvant chemotherapy, adjuvant chemotherapy, radiotherapy, endocrine therapy, etc).
3. Lymphoedema knowledge: We used Sherman's Lymphoedema Knowledge Questionnaire,[30] which consists of 20 true-false items on lymphoedema management. Each correct answer scored 1, with incorrect or unclear answers scoring 0, summing to a maximum of 20. Higher scores indicate greater knowledge. The Cronbach's alpha was 0.72.
4. Illness perception: We used the Chinese version Brief Illness Perception Questionnaire for BCRL (BIPQ-BCRL),[31] which was revised and validated based on the BIPQ by Broadbent *et al*.[32] It comprises nine items evaluating cognitive abilities, emotional responses and comprehension, with the final item focusing on etiological causality. It is scored from 0 to 10, where higher scores indicate greater negative perception. The BIPQ-BCRL demonstrated strong validity and reliability, achieving a Cronbach's alpha of 0.651 and a test–retest reliability of 0.761.[31]
5. Lymphoedema self-management behaviours: The Lymphoedema Self-Management Behaviour Questionnaire for BC, developed by Wei *et al* was used.[33] It consists of 22 items, measuring 6 self-management aspects: disease information, diet and exercise, affected limb protection, lymph reflux promotion, emotional and role management. Responses are scored on a 5-point Likert scale from 0 (never) to 4 (always), with higher scores indicating more effective self-management. The Cronbach's alpha was 0.910 and the test–retest reliability was 0.877.[33]
6. Self-efficacy: The Chinese Version of the Breast Cancer Survivor Self-Efficacy Scale (BCSES) was employed. Initially developed by Champion *et al*,[34] the Chinese version of the BCSES comprises 11 items across 2 dimensions: self-acceptance and self-development. Responses are measured using a 5-point Likert-type scale, ranging from 1 (strongly disagree) to 5 (strongly agree). The Cronbach's alpha for this version was 0.82.[35]
7. Depression and anxiety: The Hospital Anxiety and Depression Scale (HADS) was used. The HADS is a 14-item questionnaire that consists of 2 subscales of anxiety (7 items) and depression (7 items). Scores ranging from 0 to 7 indicate the absence of anxiety/depression, whereas scores ≥8 indicate the presence of anxiety/depression (8–10: borderline anxiety/depression; >10, definite anxiety/depression). It is reported to have good reliability and validity.[36]
8. Social support: The 19-item simplified Chinese version of the Medical Outcome Study Social Support Survey was used to measure social support.[37] This scale assesses four types of support: tangible, emotional/informational, positive social interaction and affectionate, using a 5-point Likert scale (1='not at all' to 5='always'). The Cronbach's α in our previous study was 0.877.[38]
9. Other factors: Other factors include exercise-related factors (regular exercise, exercise duration each day, exercise frequencies each week), accessibility to medical services, etc.

Study procedures

Cohort database construction

For this study, data collection, storage and management were conducted using the 'H6WORLD' data management platform. Prior to the start of the study, a cohort database was constructed according to the research protocol. This involved setting up the database's basic information, establishing a study-specific case report form repository, configuring form entries and variable labels, coding variables based on their characteristics, implementing logic check functions, and organising the follow-up processes as per the plan. After constructing the database, the data settings were extensively tested and refined through a pilot test. All patient data will be securely coded and encrypted with IDs to ensure data security and reliability.

Participant recruitment and baseline assessment

All eligible patients were recruited 1–3 days before surgery. Trained nurse researchers identified patients with BC with lumpectomy or mastectomy surgical plans in the medical record system and explained the cohort study to patients and their caregivers. After obtaining the informed consent, baseline assessment was performed using 'H6WORLD' data collection and management

platform with a tablet. All patients received standardised health education on lymphoedema prevention and detection. Patients and their caregivers were trained to measure limb circumference at home following the same measurement protocol. A prerecorded arm circumference measurement instruction video was also provided to them (Caregiver measurement: https://share.plvideo.cn/front/video/view?vid=0da91d76e59f1bb0d3beec630787b66c_0; Self-measurement: https://share.plvideo.cn/front/video/view?vid=0da91d76e56d72ef38a675b06fd9c188_0). All patients were provided with a follow-up package that included a BCRL health education and follow-up record manual, a retractable arm circumference tape measure, and a set of adhesive paper tapes. To ensure compliance, we established several WeChat groups for lymphoedema follow-up that patients can voluntarily join in.

Follow-up assessments

All patients will be followed up at 1, 3, 6 and 12 months postsurgery, and then semiannually until 3 years after surgery (see online supplemental table 2). Surgery-related information will be collected from medical records system at 1-month follow-up assessment. 'H6WORLD' platform will automatically send follow-up reminders via mobile messages and WeChat messages a week before follow-up time points, including links to patients' self-reported outcomes assessment forms, according to the follow-up schedule. Researchers will check completion status 1 week after the initial reminder and send another reminder to those who have not responded. If there is still no response after 1 week, patients will be contacted through phone calls. All patients can choose to undergo lymphoedema follow-up assessment at the lymphoedema follow-up centre at hospital's outpatient clinic or complete it at home. To ensure participants' adherence, we will implement the following strategies: (1) Reminder system: The data management platform automatically sends messages reminders for follow-up appointments, with additional reminders via WeChat and one-on-one phone or message reminders from researchers. Additionally, we have set up a follow-up centre in the outpatient clinic for in-person follow-up visits. (2) Incentives: Patients will receive ¥10 per completed assessment, free body composition measurements and rehabilitation materials (eg, hand grips, portable pill organisers). (3) Educational materials: A home-based self-care manual and instructional videos on measuring arm circumference are provided for home-based self-assessments. (4) Ongoing support: Continuous guidance is available through WeChat and phone consultations, offering personalised assistance with completing assessments. (5) Assessment flexibility: Follow-up assessments can be conducted either at the hospital's outpatient lymphoedema centre or completed by the patients at home.

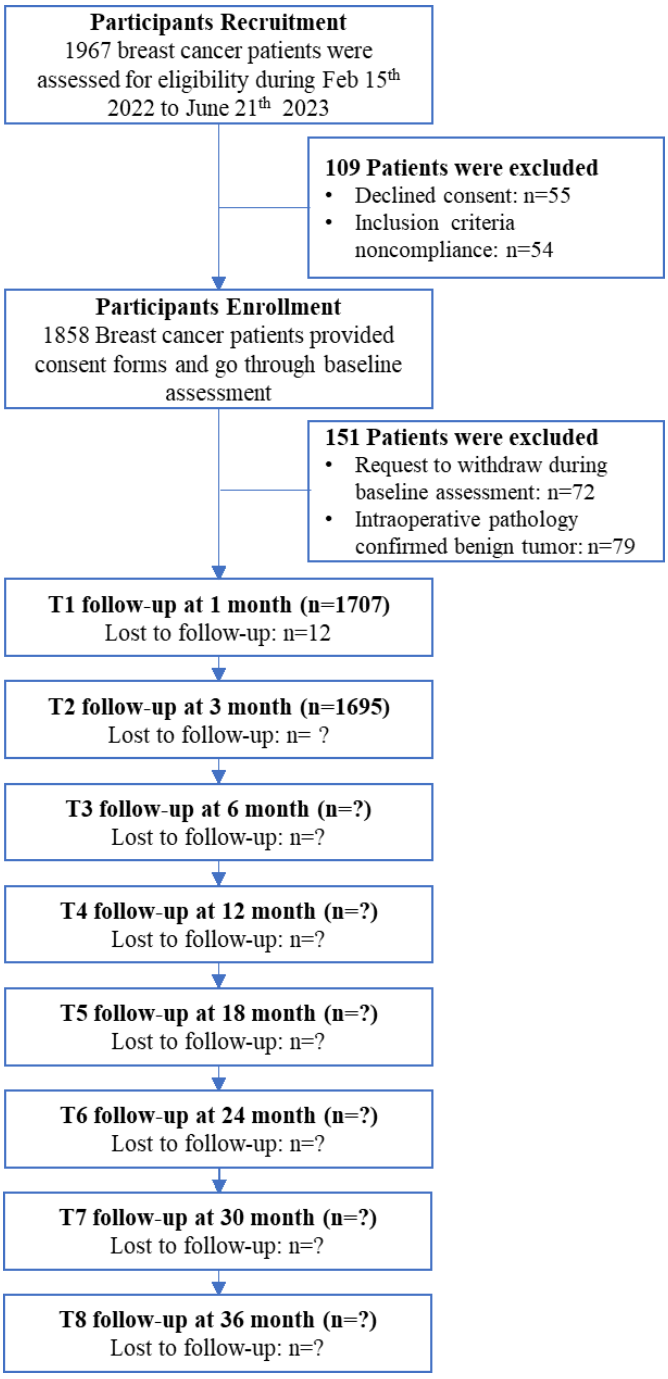

Figure 2 Research flow chart.

Data analysis

Data analysis will be performed by using SPSS V.26.0 and R V.4.2.2. When a patient develops lymphoedema, it will be considered as the endpoint event. In such cases, we will apply right-censoring by recording the time of lymphoedema onset, and these patients will no longer be included in subsequent follow-up analyses. Additionally, we will account for participants who do not complete the study due to dropout, loss to follow-up or death by applying right-censoring at their last available observation. Participants who withdraw or cannot be followed further will be censored at their last recorded assessment. For participants who pass away during

Table 1 Demographic characteristics of participants (n=1707)

Variables	Mean (SD), N (%)	Variables	N (%)
Age, years	47.6 (9.24)	Regular exercise	
Height, cm	161 (4.94)	Yes	670 (39.3)
Weight, kg	63.0 (9.48)	No	1037 (60.7)
BMI, kg/m^2	24.2 (3.44)	Daily exercise duration	
Province		<30 min/day	1071 (62.8)
Hebei	252 (14.8)	30–60 min/day	472 (27.7)
Heilongjiang	271 (15.9)	>60 min/day	163 (9.5
Inner Mongolia	183 (10.7)	Weekly exercise frequency	
Tianjin	843 (49.4)	1–2 day/week	343 (20.1)
Other	155 (9.1)	3–4 day/week	356 (20.9)
Ethnicity		At least 5 days/week	413 (24.2)
Han	1586 (92.9)	No exercise	595 (34.9)
Minority	121 (7.1)	Timeliness of medical treatment	
Marital status		Always	42 (2.5)
Single	44 (2.6)	Mostly	198 (11.6)
Married	1567 (91.8)	Sometimes	248 (14.5)
Divorced	66 (3.9)	Rarely	861 (50.4)
Widowed	30 (1.8)	Almost never	358 (21.0)
Education level		Difficulty in seeking medical care	
Primary school or below	96 (5.6)	Very easy	150 (8.8)
Middle school	483 (28.3)	Easy	378 (22.1)
High school	376 (22.1)	Average	546 (32.0)
College	295 (17.3)	Not very easy	535 (31.3)
Undergraduate	404 (23.7)	Very difficult	98 (5.7)
Postgraduate	53 (3.1)	Dominant hand with affected arm	
Residence		Yes	865 (50.7)
City	1106 (64.8)	No	842 (49.3)
Town or county	362 (21.2)	Menstrual status	
Rural or suburban area	239 (14.0)	Premenopausal	910 (53.3)
Solitary		Perimenopausal	99 (5.8)
Yes	94 (5.5)	Postmenopausal	698 (40.9)
No	1613 (94.5)	Bra cup size	
Family roles		A	243 (14.2)
Mainly cared for by others	86 (5.0)	B	700 (41.0)
Mainly caring for others	502 (29.4)	C	412 (24.1)
Caring for oneself	175 (10.3)	D	95 (5.6)
Caring for each other	944 (55.3)	E or above	12 (0.7)
Medicare payment methods		Not sure	245 (14.4)
Basic medical insurance for urban employees	1039 (60.9)	Smoking history	
Basic medical insurance for urban residents	291 (17.0)	Used to	41 (2.4)
New rural cooperative medical system	325 (19.0)	Currently (in the last month)	33 (1.9)
Commercial insurance	27 (1.6)	Never	1633 (95.7)
Public medical care	8 (0.5)	Drinking History	
Fully self-funded	17 (1.0)	Frequently	12 (0.7)
Monthly income (RMB)		Occasionally	388 (22.7)

Continued

Table 1 Continued

Variables	Mean (SD), N (%)	Variables	N (%)
<2000	255 (14.9)	Never	1307 (76.6)
2000–4000	579 (33.9)	Hormone replacement therapy history	
4000–6000	415 (24.3)	Never	1554 (91.0)
>6000	458 (26.8)	Currently undergoing	40 (2.3)
E[24] mployment		Previously undergone	113 (6.6)
Unemployed	527 (30.9)	Disease history	
Employed	666 (39.0)	None	1199 (70.2)
Retired	410 (24.0)	Hypertension	223 (13.1)
Sick leave or early retirement	15 (0.9)	Diabetes	93 (5.4)
Other	89 (5.2)	Cardiovascular disease	49 (2.9)
Occupation before surgery		Thyroid disease	131 (7.7)
Worker	212 (12.4)	Other disease history	93 (5.4)
Farmer	154 (9.0)	Family history of lymphoedema	
Company employee	348 (20.4)	Yes	17 (1.0)
Healthcare worker	70 (4.1)	No	1690 (99.0)
Education worker	125 (7.3)	Family history of breast cancer	
Business man	105 (6.2)	Yes	145 (8.5)
Administration staff	154 (9.0)	No	1562 (91.5)
Urban migrant workers	28 (1.6)	Relatives or friends diagnosed with lymphoedema	
Unemployed/housewife	397 (23.3)	Yes	78 (4.6)
Others	114 (6.7)	No	1629 (95.4)

BMI, body mass index.

the study, we will obtain information through follow-up phone calls with family caregivers, as well as medical records and death registries where available. These cases of death will be censored at the last assessment prior to death. Continuous variables, which are normally distributed or approximately normally distributed, will be summarised as means±SDs; for data not normally distributed, we will use medians and IQRs to describe. Categorical variables will be reported as frequencies and percentages. Kaplan-Meier survival curves will be used to estimate the incidence of BCRL at various postoperative follow-up intervals. Group comparisons for the presence of lymphoedema will be conducted using t-tests, χ^2 tests and non-parametric tests as appropriate to the data distribution. Missing data will be identified, reasons and mechanisms assessed and addressed using multiple imputation techniques.

Model development and internal validation

Data from the initial 1150 subjects in the cohort are used for model development and internal validation. Univariate analysis combined with COX-LASSO regression will be used to screen predictive factors, which will then be finalised in conjunction with clinical experience and expert opinion. Potential confounders, such as the number of lymph nodes removed and the type of breast surgery, will be considered. Traditional COX regression analysis and seven common survival analysis machine learning algorithms (COX, CARST, RSF, GBSM, XGBS, SSVM and SANN) will be employed for model construction and validation.[39] Bootstrap resampling will be used for internal validation, while repeated three-time fivefold cross-validation is applied for hyperparameter tuning of machine learning models. Model performance will be evaluated based on discriminative ability (eg, area under the curve) and calibration (eg, calibration curves). Clinical decision curves and model improvement indicators will be used to assess the clinical utility and predictive power of each model.[40]

Model external validation

Data from the latter 550 subjects in the cohort will be used for external validation (also temporal validation), involving the external validation and optimisation of the seven survival analysis machine learning models.[41] The validation methods are the same as those used for internal validation.

Model reporting and interpretation

After completing model validation and comparison, the recommended optimal BCRL risk prediction model will be reported and interpreted following the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis statement.[42]

Table 2 Cancer and treatment-related characteristics of participants (n=1707)

Variables	Mean (SD), median (P25, P75), n (%)	Variables	N (%)
No of LNs dissected (n=1683)	17.0 (10, 23)	Her2 level	
No of positive LNs	0 (0, 2)	Negative	916 (53.7)
Maximum tumour diameter, mm	26.3 (13.6)	Positive	321 (18.8)
Tumour laterality		Unclear	470 (27.6)
Left	890 (52.1)	Lymph vascular invasion	
Right	817 (47.9)	Yes	298 (17.5)
Tumour location		No	1222 (71.6)
Upper outer quadrant	769 (45.0)	Unclear	187 (11.0)
Lower outer quadrant	288 (16.9)	Type of breast surgery	
Upper inner quadrant	331 (19.4)	Lumpectomy	323 (18.9)
Lower inner quadrant	112 (6.6)	Mastectomy	1384 (81.1)
Areola area	108 (6.3)	Breast reconstruction	
Unclear	99 (5.8)	Tissue expander/prosthesis	102 (6.0)
Pathological grade		Autograft	12 (0.7)
I	48 (2.8)	Combined reconstruction	11 (0.6)
II	1024 (60.0)	None	1582 (92.6)
III	545 (31.9)	Surgical incision	
IV	3 (0.2)	Longitudinal incision	15 (0.9)
Unclear	87 (5.1)	Transverse incision	1265 (74.1)
Clinical tumour stage		Oblique incision	153 (9.0)
Stage 0	95 (5.6)	Crescent incision	123 (7.2)
Stage I	474 (27.8)	Laparoscopy	30 (1.8)
Stage II	848 (49.7)	Unclear	121 (7.1)
Stage III	271 (15.9)	SLNB	
Unclear	19 (1.1)	Yes	609 (35.7)
LN status		No	1098 (64.3)
Positive	561 (37.9)	ALND	
Negative	1006 (59.1)	Yes	1336 (78.3)
Unclear	52 (3.0)	No	371 (21.7)
PR status		Level of ALND	
Negative	563 (33.0)	I	227 (13.3)
Positive	1060 (62.1)	II	193 (11.3)
Unclear	84 (4.9)	III	916 (53.7)
ER status		NA	371 (21.7)
Negative	383 (22.4)		
Positive	1250 (73.2)		
Unclear	74 (4.3)		

ALND, axillary lymph nodes dissection; ER, oestrogen receptor; LN, lymph nodes; NA, not appropriate; PR, progesterone receptor; SLNB, sentinel lymph nodes biopsy.

Patient and public involvement

Patients and/or the public were not involved in the design, reporting or dissemination plans of this research. Patients with BC and their family caregivers were trained to participate in outcome measurement by self-measuring limb circumferences during long-term follow-up.

Baseline demographics

Study recruitment

From 15 February 2022 to 21 June 2023, a total of 1967 patients with BC were assessed for eligibility. Of these, 109 were excluded: 55 declined consents and 54 did not meet inclusion criteria. Thus, 1858 patients consented

and underwent baseline assessment, but 151 were further excluded: 72 withdrew and 79 had benign tumours. This left 1707 patients for follow-up. At the 1-month follow-up (T1), 12 were lost, leaving 1695 participants. Further updates will include detailed follow-up results for each phase (see figure 2).

Demographic characteristics

Participants were drawn from 20 provinces across China, with the majority coming from Tianjin, Heilongjiang, Hebei and Inner Mongolia. The mean age of 1707 participants was 47.6 years (SD=9.24, range: 22–75), with an average BMI of 24.2 kg/m² (SD=3.44, range: 15.0–37.1) (see online supplemental figure 2 for age and BMI distribution). The majority (60.8%) of them are aged from 40 to 55 years. Most participants were married (91.8%). The majority of them (66.1%) had an education level above middle school and lived in cities (64.8%). A majority (94.5%) did not live alone, and 55.3% provided mutual family care. Insurance coverage was predominantly basic medical insurance for urban employees (60.9%). In terms of lifestyle, 39.3% exercised regularly, primarily under 30 min daily (62.8%). Employment status showed 39.0% employed. Most of them never smoked (95.7%) and never drank alcohol (76.6%). 13.1% had the diagnosis of hypertension. 8.5% of them had a family history of BC and 4.6% had relatives or friends with lymphoedema (see table 1).

Disease and treatment-related characteristics

The average maximum tumour diameter was 26.3 mm (SD=13.6). Tumours were almost evenly split between the left (52.1%) and right (47.9%) breasts, most commonly located in the upper outer quadrant (45.0%). The majority of tumours were grade II (60.0%) and stage II (49.7%). Mastectomy was performed in 81.1% of cases, with 18.9% undergoing lumpectomy. SLNB was conducted in 35.7%, while ALND was performed in 78.3%. The median number of lymph nodes dissected was 17 (IQR: 10–23). Transverse incisions were most common (74.1%), and lymph vascular invasion was noted in 17.5% of participants (see table 2).

Psychosocial-behavioural characteristics

Participants had a mean lymphoedema knowledge score of 14.3 (SD=3.70) and self-efficacy of 31.2 (SD=6.14), with self-acceptance and self-development scores of 13.9 (SD=3.07) and 17.3 (SD=3.63), respectively. The overall social support score averaged 74.8 (SD=14.5). Lymphoedema illness perception scored 55.4 (SD=13.3). Mean scores for anxiety and depression were 6.5 (SD=3.85) and 4.8 (SD=3.76), respectively. The majority of the participants were without symptoms of anxiety (65.8%) and depression (77.3%) (see online supplemental table 3).

Presurgery limb circumferences and volumes

The presurgery limb circumferences and volumes are shown in online supplemental table 4. The mean presurgery volume was 1875.51 (SD=333.36) mL for the left arm and 1893.43 (SD=331.37) mL for the right arm.

ETHICS AND DISSEMINATION

The study protocol has been approved by the Biomedical Ethics Committee of Peking University (IRB00001052-21124) and the Research Ethics Committee of Tianjin Medical University Cancer Institute and Hospital (bc2023013). All participants were required to provide written informed consent. This cohort study was totally voluntary, anonymous and confidential. The study results will be published in peer-reviewed journals and will be presented at several research conferences.

Author affiliations
[1]Department of Nursing, Tianjin Medical University Cancer Institute & Hospital, National Clinical Research Center for Cancer, Tianjin, China
[2]Peking University School of Nursing, Beijing, China
[3]Department of Thyroid and Breast Surgery, Peking University First Hospital, Beijing, China
[4]Department of General Surgery, Peking University Third Hospital, Beijing, China
[5]The First Department of Breast Surgery, Tianjin Medical University Cancer Institute & Hospital, National Clinical Research Center for Cancer, Tianjin, China
[6]Department of Cancer Epidemiology and Biostatistics, Tianjin Medical University Cancer Institute & Hospital, National Clinical Research Center for Cancer, Tianjin, China
[7]Department of Nursing, Peking University Third Hospital, Beijing, China
[8]School of Nursing, Tianjin Medical University, Tianjin, China
[9]Lymphedema Clinic, Tianjin Medical University Cancer Institute & Hospital, Tianjin, China

Acknowledgements The authors were grateful to all the patients who participated in this study. Thanks to all nurses and doctors in Tianjin Medical University Cancer Institute & Hospital for their assistance in participant recruitment and follow-up. Special thanks to the support provided by the China Scholarship Council and the Center for Global Initiatives at Johns Hopkins University School of Nursing.

Contributors All authors contributed to the development of the study. AS and QL led the conception and design of the study in consultation with YZ, JY, YueW, WQ, HongmeiZ and PW. AS, XL, ZZ, JB, LZ, PW, YingW and WQ recruited participants and assisted with the follow-up visits. HongmengZ reviewed the manuscript from the perspective of a breast cancer physician and YH revised the manuscript as an epidemiological statistician. AS wrote and submitted the manuscript. All authors reviewed and approved the final manuscript. QL acted as the guarantor.

Funding This study was supported by grants from the National Natural Science Foundation of China (72174011), Nursing Innovation Talent Fund of Tianjin Medical University Cancer Institute and Hospital (HL2021-27), and Tianjin Key Medical Discipline (Specialty) Construction Project (TJYXZDXK 011A).

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Consent obtained directly from patient(s).

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iDs
Aomei Shen http://orcid.org/0000-0003-0569-8746
Qian Lu http://orcid.org/0000-0003-2611-3284

REFERENCES

1 Shen A, Lu Q, Fu X, *et al*. Risk factors of unilateral breast cancer-related lymphedema: an updated systematic review and meta-analysis of 84 cohort studies. *Support Care Cancer* 2022;31:18.
2 Ahn HR, Jeong HE, Jeong C, *et al*. Incidence and risk factors of breast cancer-related lymphedema in Korea: a nationwide retrospective cohort study. *Int J Surg* 2024;110:3518–26.
3 McLaughlin SA, Wright MJ, Morris KT, *et al*. Prevalence of lymphedema in women with breast cancer 5 years after sentinel lymph node biopsy or axillary dissection: patient perceptions and precautionary behaviors. *J Clin Oncol* 2008;26:5220–6.
4 Togawa K, Ma H, Smith AW, *et al*. Self-reported symptoms of arm lymphedema and health-related quality of life among female breast cancer survivors. *Sci Rep* 2021;11:10701.
5 Bian J, Shen A, Yang W, *et al*. Financial toxicity experienced by patients with breast cancer-related lymphedema: a systematic review. *Support Care Cancer* 2023;31:354.
6 Eaton LH, Narkthong N, Hulett JM. Psychosocial Issues Associated with Breast Cancer-Related Lymphedema: a Literature Review. *Curr Breast Cancer Rep* 2020;12:216–24.
7 Gursen C, Dylke ES, Moloney N, *et al*. Self-reported signs and symptoms of secondary upper limb lymphoedema related to breast cancer treatment: Systematic review. *Eur J Cancer Care* 2021;30:1–22.
8 Vignes S, Fau-Prudhomot P, Simon L, *et al*. Impact of breast cancer-related lymphedema on working women. *Support Care Cancer* 2020;28:79–85.
9 Jørgensen MG, Toyserkani NM, Hansen FG, *et al*. The impact of lymphedema on health-related quality of life up to 10 years after breast cancer treatment. *NPJ Breast Cancer* 2021;7:70.
10 Shen A, Qiang W, Zhang L, *et al*. Risk Factors for Breast Cancer-Related Lymphedema: An Umbrella Review. *Ann Surg Oncol* 2024;31:284–302.
11 Shen A, Lu Q, Zhang L, *et al*. Risk factors of breast cancer-related lymphoedema: protocol of an umbrella review. *BMJ Open* 2023;13:e070907.
12 Clough-Gorr KM, Ganz PA, Silliman RA. Older breast cancer survivors: factors associated with self-reported symptoms of persistent lymphedema over 7 years of follow-up. *Breast J* 2010;16:147–55.
13 Paskett ED, Naughton MJ, McCoy TP, *et al*. The epidemiology of arm and hand swelling in premenopausal breast cancer survivors. *Cancer Epidemiol Biomarkers Prev* 2007;16:775–82.
14 Lin Q, Yang T, Yongmei J, *et al*. Prediction models for breast cancer-related lymphedema: a systematic review and critical appraisal. *Syst Rev* 2022;11:217.
15 Shen A, Wei X, Zhu F, *et al*. Risk prediction models for breast cancer-related lymphedema: A systematic review and meta-analysis. *Eur J Oncol Nurs* 2023;64:S1462-3889(23)00060-1.
16 Koelmeyer LA, Gaitatzis K, Dietrich MS, *et al*. Risk factors for breast cancer-related lymphedema in patients undergoing 3 years of prospective surveillance with intervention. *Cancer* 2022;128:3408–15.
17 Grzywacz JG, Fuqua J. The social ecology of health: leverage points and linkages. *Behav Med* 2000;26:101–15.
18 Riley RD, Ensor J, Snell KIE, *et al*. Calculating the sample size required for developing a clinical prediction model. *BMJ* 2020;368:m441.
19 Riley RD, Snell KI, Ensor J, *et al*. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. *Stat Med* 2019;38:1276–96.
20 Vergouwe Y, Steyerberg EW, Eijkemans MJC, *et al*. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *J Clin Epidemiol* 2005;58:475–83.
21 Brunelle CL, Taghian AG. Breast Cancer–Related Lymphedema: the Prospective Surveillance Model, Early Intervention Strategies, and Role of Complete Decongestive Therapy. *Curr Breast Cancer Rep* 2023;15:12–23.
22 Levenhagen K, Davies C, Perdomo M, *et al*. Diagnosis of Upper Quadrant Lymphedema Secondary to Cancer: Clinical Practice Guideline From the Oncology Section of the American Physical Therapy Association. *Phys Ther* 2017;97:729–45.
23 Brunelle CL, Roberts SA, Horick NK, *et al*. Integrating Symptoms Into the Diagnostic Criteria for Breast Cancer-Related Lymphedema: Applying Results From a Prospective Surveillance Program. *Phys Ther* 2020;100:2186–97.
24 Shi S, Lu Q, Fu MR, *et al*. Psychometric properties of the Breast Cancer and Lymphedema Symptom Experience Index: The Chinese version. *Eur J Oncol Nurs* 2016;20:10–6.
25 Dylke E. Measurement of breast cancer-related lymphoedema. *J Physiother* 2022;68:238–43.
26 Rafn BS, McNeely ML, Camp PG, *et al*. Self-Measured Arm Circumference in Women With Breast Cancer Is Reliable and Valid. *Phys Ther* 2019;99:240–53.
27 Rafn BS, Singh CA, Midtgaard J, *et al*. Self-Managed Surveillance for Breast Cancer-Related Upper Body Issues: A Feasibility and Reliability Study. *Phys Ther* 2020;100:468–76.
28 Karges JR, Mark BE, Stikeleather SJ, *et al*. Concurrent Validity of Upper-Extremity Volume Estimates: Comparison of Calculated Volume Derived From Girth Measurements and Water Displacement Volume. *Phys Ther* 2003;83:134–45.
29 Ancukiewicz M, Russell TA, Otoole J, *et al*. Standardized method for quantification of developing lymphedema in patients treated for breast cancer. *Int J Radiat Oncol Biol Phys* 2011;79:1436–43.
30 Sherman KA, Koelmeyer L. The role of information sources and objective risk status on lymphedema risk-minimization behaviors in women recently diagnosed with breast cancer. *Oncol Nurs Forum* 2011;38:E27–36.
31 Shen A, Zhu F, Zhang L, *et al*. Test of Reliability and Validity of the Brief Illness Perception Questionnaire for Breast Cancer-Related Lymphedema. *J Nurs* 2023;30:54–9.
32 Broadbent E, Petrie KJ, Main J, *et al*. The brief illness perception questionnaire. *J Psychosom Res* 2006;60:631–7.
33 Wei X, Lu Q, Shen A, *et al*. Development and psychometric evaluation of a lymphoedema self-management behaviour questionnaire for breast cancer patients. *Eur J Cancer Care (Engl)* 2022;31:e13743.
34 Champion VL, Ziner KW, Monahan PO, *et al*. Development and psychometric testing of a breast cancer survivor self-efficacy scale. *Oncol Nurs Forum* 2013;40:E403–10.
35 Yuan X, Wu F, Howell D, *et al*. Translation, Validity, and Reliability of the Chinese Version of the Breast Cancer Survivor Self-efficacy Scale in China. *Cancer Nurs* 2019;42:E31–40.
36 Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand* 1983;67:361–70.
37 Sherbourne CD, Stewart AL. The MOS social support survey. *Soc Sci Med* 1991;32:705–14.
38 Shen A, Wu P, Qiang W, *et al*. Factors associated with lymphedema self-management behaviours among breast cancer survivors: A cross-sectional study. *J Clin Nurs* 2023;32:7330–45.
39 Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. *Science* 2015;349:255–60.
40 Steyerberg EW. *Clinical prediction models a practical approach to development, validation, and updating*. New York: Springer, 2019.
41 Altman DG, Vergouwe Y, Royston P, *et al*. Prognosis and prognostic research: validating a prognostic model. *BMJ* 2009;338:bmj.b605.
42 Collins GS, Reitsma JB, Altman DG, *et al*. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* 2015;350:g7594.