

BMJ Open Patterns of cancer screening, incidence and treatment disparities in China: protocol for a population-based study

Nengliang Yao,¹ Jialin Wang,² Yuanchu Cai,³ Jing Yuan,⁴ Haipeng Wang,³ Jiyong Gong,² Roger Anderson,¹ Xiaojie Sun³

To cite: Yao N, Wang J, Cai Y, *et al.* Patterns of cancer screening, incidence and treatment disparities in China: protocol for a population-based study. *BMJ Open* 2016;**6**:e012028. doi:10.1136/bmjopen-2016-012028

► Prepublication history for this paper is available online. To view these files please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2016-012028>).

Received 23 March 2016

Revised 24 June 2016

Accepted 18 July 2016



CrossMark

For numbered affiliations see end of article.

Correspondence to

Professor Xiaojie Sun;
xiaojiesun@sdu.edu.cn
and

Professor Jialin Wang;
wangjialin6681@sina.com

ABSTRACT

Introduction: Cancer has become the leading cause of death in China. Several knowledge gaps exist with respect to the patterns of cancer care and disparities in China. Chinese healthcare researchers do not have access to cancer research data of high quality. Only cancer incidence and mortality rates have been analysed in China while the patterns of cancer screening and treatment and disparities have not been rigorously examined. Potential disparities in cancer care by socioeconomic status have not been analysed in the previous literature. Population-based estimates of cancer care costs remain unexamined in China. This project will depict the pattern of cancer screening, incidence and treatment in Shandong province and enhance our understanding of causes of disparities in cancer control.

Methods and analysis: We will create the first linked database of cancer registry and health insurance claims in China. We obtained cancer registry data on breast, gastrointestinal and lung cancer incidence from 2011 to 2014 and their health insurance claims information from 6 cities/counties of 10.63 million population and validated it with hospital discharge data. A 1600 participant survey will be administered to collect additional information of patients' socioeconomic status, employment and cancer care costs. Frequency analysis, spatial data exploratory analysis, multivariate logistic regression with instrumental variable, generalised linear regression and subgroup analysis will be used to analyse the following: the receipt of cancer screening, stage at diagnosis, guideline-concordant treatment and cancer care costs. Patient characteristics, tumour features, hospital characteristics, patient comorbidities and county-level descriptors will be used as covariates in the multivariate analysis.

Ethics and dissemination: The Institutional Review Board of the School of Public Health of Shandong University approved this study (20140201). Data compiled from this project will be made available to all Chinese healthcare researchers. Study results will be disseminated through peer-reviewed publications and presentations at national and international meetings.

INTRODUCTION

Cancer incidence and mortality are very high in China. China is home to about a fifth of

Strengths and limitations of this study

- We will create the first linked data of cancer registry and insurance claims in China to conduct health services and policy research.
- We will identify innovative scientific opportunities to improve cancer control and reduce inequities in communities experiencing an excess burden of cancer.
- We will disseminate the findings to communities, policymakers, healthcare providers and the scientific community.
- It is possible that unobservable confounding factors substantially influence the study outcomes.
- We are not able to study clinical efficacy such as survival or mortality as the outcome variable in the project due to incomplete records of death in China.

the world's population, but it accounts for 27% of global cancer deaths.¹ Cancer has become the leading cause of death in China (374.1 per 100 000 person-years).² It is also the most feared disease although many types of cancer are curable. Both incidence of cancer and mortality due to it in China have been increasing.³ The most recent data show that 13% of deaths in China were caused by malignant neoplasm and every minute, six people in China are diagnosed with cancer.³

Several knowledge gaps exist with respect to the disparities in cancer control in China. First, only incidence of cancer and mortality rates have been analysed in China, while the patterns of cancer screening and treatment and disparities have not been rigorously examined. For example, the most recent data show that cancer death rates in the urban areas have decreased in the past few decades; however, rural areas with fewer cancer care resources have not experienced comparable decline in mortality rates.³ Despite its excess, mortality due to cancer and its apparent variation in economic

prosperity, population density and low proximity of cancer care, presently little is known about cancer treatment patterns in rural China. Second, potential disparities in incidence of cancer or mortality because of patient's family structure, marital status, occupation, insurance coverage and income have not been analysed in the earlier literature. For example, little research has been carried out to analyse the effect of marriage on elevated cancer mortality in Chinese men. Third, cancer care costs remain unexamined in China from the patient's perspective. For example, patients with cancer were more likely to go bankrupt in the USA,⁴ but few researchers have examined the direct and indirect cost of cancer care in China. Finally, although sex, rural residence and geographic region have been found to be correlated with disparities in incidence of cancer and mortality, no rigorous analysis has examined these factors adjusting for other important patient, tumour and hospital characteristics.

Lung, female breast and gastrointestinal cancers are of particular interest in China because they are among the most prevalent types of cancer, require a coordinated and multidisciplinary approach to treatment and have well-documented treatment guidelines and outcomes. The most common types of cancer in China appear to be that of the lungs, breasts and the gastrointestinal tract. Lung cancer has increased almost five times during the past three decades and has become the leading cause of death.⁵ Breast cancer among women has doubled in the past 30 years.⁶ Digestive tract cancers continued to stay among top three cancers during the past three decades.⁵ Thanks to the development of healthcare technology, medical screening can detect breast, colorectal and lung cancers before the symptoms start to show.⁷⁻⁹ Treatment guidelines for these three cancers have been created and updated periodically to promote continuous quality improvement.¹⁰⁻¹² These screening and treatment guidelines are well recognised and have influence outside as well as inside the USA because there is a causal link between the delivery of guideline-concordant care and a substantial survival benefit. Presently little is known about the guideline concordance of cancer screening and treatment in China.

Linked data of cancer registry, insurance claims and hospital discharge information provide a unique population-based source of information to study cancer care outcomes. The data from cancer registries include clinical, demographic and cause of death information for persons with cancer. The insurance claims have information for healthcare services from insurance enrolment until death. Hospital discharge data provide information on any discharge from hospitals in Shandong province. The linkage of these three data sources can be used for an array of epidemiological and health services research. Numerous cancer care studies have been carried out in the USA using similarly linked data.¹³ These published studies have contributed important literature to cancer care practice

and policy. For example, investigators used this combined data set to study patterns of care for persons with cancer before a cancer diagnosis, over the period of initial diagnosis and treatment and during long-term follow-up. Investigators have also examined the comparative effectiveness of different tests and procedures and the costs of cancer treatment. Presently no published study has used this type of linked data to study cancer care in China.

High-quality population-based cancer incidence data have been collected since the early 1980s in China and published periodically in well-regarded journals.¹⁴⁻²⁰ In addition, the data completeness seems not to differ between rural and urban areas. A recent paper has found that the rates of under-reporting are 3.6% in a rural area of Hebei province during 2002-2007.²¹ Another study in Shenzhen city has found the rates of under-reporting are 3.7% during 2000-2009.²² Furthermore, the rates of under-reporting of lung, female breast and gastrointestinal cancers are considered lower than other cancer types because of its prevalence.²³ For this particular project, we will select five counties from 17 pilot counties in Shandong province that participated in a nationwide programme which aims for early diagnosis and early treatment of patients with cancer. The registry data in these pilot counties have better data quality than the national average. In summary, we expect that 97-99% of patients with cancer are in the current registries.

Regarding the insurance rates, China's leadership had taken bold steps to accelerate improvement, including increasing government spending on health and committing to reach 100% insurance coverage by 2010.²⁴ The Minister of Health, Dr Chen Zhu, has announced that 96% of Chinese people are insured at the end of 2011,²⁵ which is comparable to the insurance rates of US citizens 65 years and older. As an eastern coastal province in China, Shandong had higher insurance rates than the national average. Many health service research studies have been carried out by using the Chinese health insurance claims data.²⁶⁻²⁸ Hospitals provide both outpatient and inpatient care to patients with cancer. The outpatient and inpatient records are all stored in the electronic record system at each hospital. So we expect that a vast majority of patients with cancer in the registry data had claims data and hospital records.

This project will identify innovative scientific opportunities to improve cancer control and reduce inequities in communities experiencing an excess burden of cancer. We will disseminate findings to communities, policymakers, healthcare providers and the scientific community. The major findings of this project will be submitted to the National Health and Family Planning Commission of China. We will propose policy recommendations to reduce the cancer burden in China. Findings from the proposed study could be valuable in assisting community cancer care agencies and planners to address gaps in cancer care. We will also train

students and young investigators to be part of the next generation of cancer care researchers. In addition, this project will create state-of-the-art data that will reflect the linkage of cancer registry file, health insurance claims and hospital discharge information. Finally, our findings will contribute towards the cumulative knowledge of cancer care globally.

AIMS AND OBJECTIVES

This project has four goals:

- ▶ To examine patterns in stage at diagnosis, screening and treatment. (A) Are there disproportionate late stage cancers in specific population groups by patient characteristics? (B) Are there differences in cancer screening rates across populations by demographic characteristics and socioeconomic status? (C) Are recommended cancer treatments underused in some subgroups of the population in China?
- ▶ To examine associations among the patient, tumour and hospital characteristics and cancer disparities in stages at diagnosis, screening and recommended treatment. (A) What factors are related to these disparities in stages at diagnosis, screening and treatment? (B) Are there any causal relationships? (C) Are these causal factors modifiable?
- ▶ To estimate the change and disparities in total costs as a function of changes in practice patterns for the treatment of lung, female breast and gastrointestinal cancer. To examine how practice characteristics have influenced treatment costs. Did patients with cancer remain employed? How does it differ across population groups?
- ▶ To disseminate findings to policymakers, communities, healthcare providers and the scientific community.

METHODS AND ANALYSIS

A conceptual framework for patient access to high-quality care

The conceptual model is presented in [figure 1](#).²⁹ The patient's predisposing factors (such as health insurance, age, socioeconomic status, comorbidity and clinical eligibility for treatment) interact with systems factors such as the accessibility or supply of cancer care services (determined by the patient's location) to determine the patient's choice of cancer care providers. In turn, the likelihood of optimal treatment is influenced by the characteristics of the providers who are selected to treat the patient. The latter oncology resources available to each patient can be characterised in terms of a few key features hypothesised to promote quality of care, including (1) volume of cancer care provided by facilities and surgeons; (2) comprehensiveness of services, so that each recommended component of care is available from a single organisational entity, without referral to other providers; and (3) proximity of these types of providers to the patient. In other words, quality of care for any individual patient is, to a measurable degree, influenced

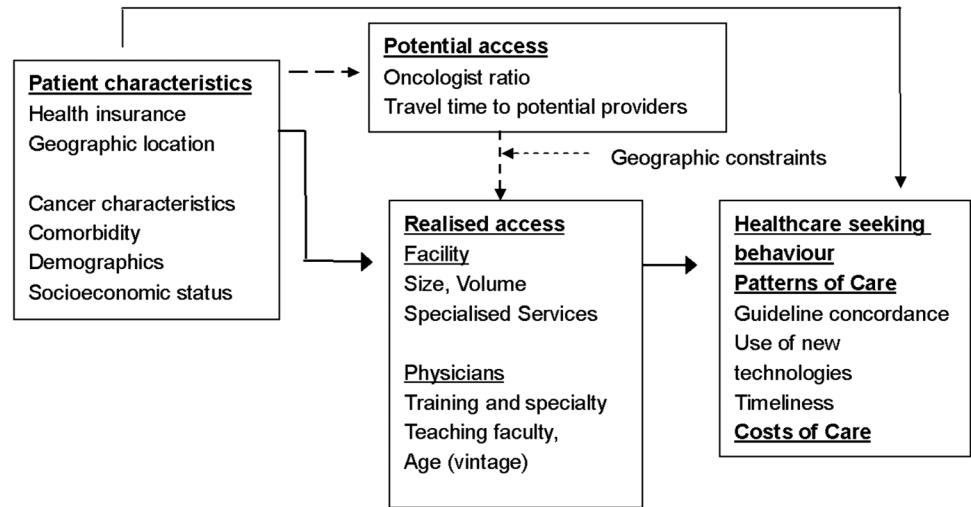
by the translation of potential access into realised access. While oversimplified, our model helps organise and guide our analysis of the determinants of variation in cancer screening and treatment.

Patient characteristics: Patient characteristics are modelled in [figure 1](#) as having both direct and indirect effects on treatment. Among direct effects, patient characteristics such as age and tolerability (comorbidity) are by far the largest influences on treatment pattern.³⁰ Age is uniquely important for treatment patterns, apart from the risk of concurrent health conditions. For colon cancer, guidelines include age 80 years and therefore are less age constrained, and major therapies have been shown to not be more toxic with increased age.

Geographic location may constrain realised access: Patients' geographic location may affect treatment by constraining access to centres or facilities identified to have a certain propensity for delivering guideline-concordant care either to a subgroup of patients to which the specific patient belongs, or more generally to the population of patients treated. Modelling geographic access effects of patient characteristics, as shown in [figure 1](#), recognises that healthcare supply and proximity to patient need may be unbalanced. Physicians do not preferentially seek to practice in regions where supply is shortest, such as rural areas, but instead, may choose affluent or more densely populated locations that yield an economical patient mix to buffer cost reimbursement constraints and achieve a desirable patient load.³¹ Since we theorise that patients prefer access to the nearest cancer treatment facility for their cancer treatment (based on community ties, time savings and overall convenience), patients geographic location and distance (or travel time) to various types of cancer treatment providers (potential access) will constrain actual or 'realised access'. Realised access to care in areas with ample physician supply is influenced by admission policies, provider and facility type and nearness.³² In areas with low supply, access is heavily constrained and favours realised access to the nearest treatment facility.

Close, comprehensive treatment facilities of this type should reduce concordance with guidelines by promoting access to an alternate, nearer and/or low-volume facilities. Studies relating distance to cancer treatment providers with patterns of cancer care have tended to confirm that distance is a barrier that negatively affects care. An interaction of the effect of distance on radiation use with the region of the country and nodal status may also present. Our conceptual model 2 also addresses a methodological complication of 'selective referral' in associations of provider type and distance on patient outcomes. Selective referral refers to the possibility that patients who travel greater distances to large, comprehensive cancer centres will differ along unmeasured dimensions (involving clinical status or the patient's personality and sophistication) that also affect concordance with standard treatment guidelines.

Figure 1 A conceptual model of determinants of patterns of care for patients with cancer.



Since there is considerable variation in local cancer care supply and distance (or travel time) to nearest comprehensive cancer care facility, our study of patterns of care can contribute to a better understanding of the effects of provider type and distance on concordance with screening and treatment guidelines, while examining these issues will also inform efforts to improve the quality of care for patients living in communities with elevated cancer burden. We expect that many patients in Shandong have no nearby oncology services and have to travel considerable distances to the nearest providers able to provide the care recommended by the guidelines; some patients could potentially access local cancer treatment providers, but opt to travel to distant comprehensive centres; still others live near and routinely access comprehensive centres. Owing to these variations in screening and treatment pattern, we also expect significant variation in cancer care costs across patient subgroups by patient and community characteristics.

Selection of cases

Seventeen counties in Shandong participated in a large nationwide programme aims for early diagnosis and early treatment of patients with cancer. JW is the principal investigator of Shandong subproject. Five representative counties will be drawn from these 17 counties. Three of five urban districts of Ji'nan city will also be included to represent the metropolitan area in this project. Cases will be drawn from Ji'nan city and five counties of Shandong province. Patients include all confirmed cases of primary lung and bronchus (C34.0–C34.9), female breast (C50.0–C50.9) and gastrointestinal cancers (especially colorectal cancers (C18.0–C18.9; C19.9–C20.9)) identified as diagnosed in 2011, 2012, 2013 and 2014 calendar incidence years (rural cases: 2011–2013; urban cases: 2011–2014). Included cases should be microscopically confirmed. Cases reported from autopsy or death certificate will be excluded. Breast cancer cases coded with Paget's disease for breast cancer, mesotheliomas, Kaposi sarcoma, lymphohaematopoietic malignancies and lobular carcinoma

will be excluded. Colon cancer cases will be excluded which include histologies other than adenocarcinoma (ie, lymphomas, sarcomas, squamous cell carcinomas, carcinoids). Finally small cell lung cancer will be excluded in this study.

Registry data procedures

We will obtain central cancer registry data of lung, female breast and gastrointestinal cancers diagnosed and reported for the calendar year 2011–2014 (rural data: 2011–2013; urban data: 2011–2014). All consolidated data will be uploaded into EXCEL software. We will review the records from each county/district for data completeness, eligibility for study (class of case codes and text notes). For cases with missing stage or other key clinical status data, our registry data quality expert will inspect the individual abstracts for the case to review each facility codes and text.

Screening and treatment information

Cancer programmes do not routinely collect central cancer registry screening and treatment data and thus, they are not adequate for research. We will use health insurance claims and hospital discharge data to track the screening and treatment of patients. For each study case, we will request all insurance claims and hospital discharge data available for a minimum period 12 months postdiagnosis.

Potential conflict information

We will identify the treatment information from the linked data and compare with the clinical guidelines. Since the data structure is similar to the registry claims linked cancer data in the USA, we are confident that a decision of guideline concordance can be made. If there is conflicting treatment information in the linked data, we will use hospital records and/or claims data to decide if the treatment is concordant with the clinical guidelines, because hospital and claims data are considered more accurate than registry data. The importance

of this proposed project is that it will help establish the current practical uses of China's cancer registry data system and document where improvement is needed. Initially, with the current data system, we expect to report on the patterns of care of treated patients—where data on the first course of treatment is present. We will focus on locoregional treatment as the initial treatment patterns as these data are collected from a single treatment facility.

Linking data and evaluating the match

Health insurance claims, hospital discharge information and consolidated cancer registry data will be merged with a probabilistic match algorithm. The standard matching string consists of a national ID number and last name. We will review non-matches for cases, and if substantial we will request that non-matches are passed to a second match string consisting of full name, date of birth, sex and county of residence.

Physician administrated patient survey

A survey questionnaire assists in collecting additional information of indirect costs that are not documented in the aforementioned linked data. Chinese researchers have used the survey method to analyse chemotherapy and complementary medicine for patients with cancer.^{33 34} The survey is often administrated in the cancer treatment facilities or the homes of the patients with cancer. Public health staff, selected from local centres for disease prevention and control (CDCs), are responsible for introducing the study to all potential participants and then determining their eligibility for inclusion. In the present study, we will use a modified version of a questionnaire developed by the Penn State Cancer Survivor Survey and Virginia Commonwealth University Cancer Survivor Survey.^{35 36}

A randomly selected group of patients will be surveyed by the local public health staff in the household survey. The selected patient should be: (1) aged 18 or older; (2) had a lung, female breast or gastrointestinal cancer diagnosis; (3) had undergone primary cancer treatment; (4) fluency in oral Mandarin; and (5) mentally and physically competent for the survey in the opinion of the supervising physician. Patients will be screened for participation in the field work by reviewing their electronic medical records (inpatients) or in consultation with the charge nurse (outpatients). Potential participants will be approached and asked simple screening questions. Eligible participants will be provided informed consent before the survey. Participants are required to complete the questionnaire in a single session. Further explanation can be offered by the investigator when a participant reported difficulty in understanding any item. For participants with low literacy, impaired vision or poor health status, the physician will read each question and response option to the patient item by item and record his or her response.

Sample size

For the primary study aims, an estimated total of 7500 cases of lung cancer, 7200 cases of female breast cancer and 4300 gastrointestinal cancers cases will be identified from the registry files which meet the study criteria and expected to link to claims and discharge data. Sample size estimates are derived from actual incidence rates and population size of the study region. For the patient survey, around 1600 cases (1000 rural cases and 600 urban cases) will be surveyed. We expect to see a response rate close to 90%. The assessment of sample size for the survey investigation is based on available manpower and ensuring adequate statistical power to detect differences in means between the rural and urban patients. We anticipate a sample size 1000 rural patients and 600 urban patients during this period. We will test the null hypothesis using a two-sided test at significance level $\alpha=0.05$ with at least 80% power in γ regression in the PASS V.12 software package.

Specification of primary outcome variables

We will use treatment guidelines published in 2010 and screening guidelines published before 2010 as the basis for our 'guidelines for care'. The primary outcomes include (1) incidence rates, (2) proportion of late stage diagnosis, (3) concordance with screening and treatment recommendations and (4) cancer care costs. Incidence rates and proportion of late stage diagnosis will be computed with the readily available information in the registry file. The direct cancer care costs include the payment by public insurance, employer's subsidy and the patient's out-of-pocket expense that are available in the linked data. The indirect cancer care costs include travel, absenteeism, reduced or lost productivity, and unemployment of the patient and the caregiver. Guideline concordance will be computed with the linked data. The National Comprehensive Cancer Network (NCCN) and the US Preventative Services Task Force (USPSTF) have produced and regularly updates a comprehensive set of evidence-based cancer screening and treatment guidelines and serves as a basis in numerous patterns of care studies. For each cancer site, we will assess practice performance by coding each definitive guideline recommendation used to define quality measures for lung, breast and gastrointestinal cancer and determine concordance with each measure among those eligible to receive that screening and treatment. These measures were selected based on their impact on disease-free and overall survival (OS), the degree to which opportunities for improvement exist, and the feasibility of data collection. For this study, 'initial therapy' is defined as therapy that is provided within 12 months of diagnosis and which does not occur after a second restaging, as the latter would indicate a high likelihood of early recurrence.

Potential long-term outcomes

Although we are limited by the available funding and follow-up time, we will explore on feasibility and testing basis linking death data to the cancer registry in the

third year of this project. Our currently proposed study is developmental in that we will for the first time assess data quality, reasonable use and areas for improvement modelled against the US Surveillance, Epidemiology, and End Results (SEER)-based systems.

Clinical and treatment variables

Staging

In collaborative staging, registrars code required tumour information relevant to tumour size, extension, lymph nodes (evaluated and positive), site-specific factors and distant metastasis and rely on computer-generated values for the American Joint Committee on Cancer (AJCC) T, N, M and Stage Group.

Surgical approach

Claims and discharge data provide valuable information not found in registry regarding the intent and purpose of procedures. Date of surgical procedure for the primary site will be used to determine when surgery was given as treatment for cancer, and scope of lymph node surgery will be used to document removal, biopsy or aspiration of regional lymph node(s) at the time of surgery of the primary site or during a separate surgical event.

Neoadjuvant and adjuvant radiation therapy

Radiation therapy and date started and ended from discharge or claims data, whichever provides positive evidence of radiotherapy received: cases where neither claims nor discharge data indicate radiotherapy was given will be classified as 'no radiation'. Cases where either data source lists radiation therapy given will be classified as receiving radiation therapy.

Systemic therapy

Claims and discharge data will be used to indicate planning of chemotherapy as a single agent, multiple agents, not planned and planned but not given. Cases where neither claims nor discharge data indicate that chemotherapy services were provided will be listed as 'no chemo'.

Hormonal therapies

Treatment with tamoxifen, arimidex, aromasin, letrozole, exemestane, anastrozole, fulvestrant, goserelin, leuprolide and megestrol acetate are prescription drugs and will be located within both discharge and claims files.

Direct cancer care costs

All health insurance claims files will be used to estimate the costs of care of patients with cancer and non-cancer control individuals. We will use the sum of insurance payments, subsidy and out-of-pocket expenses to reflect costs of care. All cost estimates will be reported in 2014 Chinese yuan (can be converted into US dollars by dividing by 6.3–6.5). Within each phase of care and for

each tumour site, we will calculate the total costs of care and months of observation of patients with cancer and control individuals. The mean net monthly cost by the phase of care was estimated as the difference in cost for patients with cancer and non-cancer control individuals. CIs will be calculated by use of the large sample normal approximation to the mean. We will also evaluate the net costs of care by stage at diagnosis in the initial.

Indirect cancer care costs

We will develop interviews and related cost estimation algorithms for measuring costs that result from cancer care for the patient and a caregiver. We defined indirect costs as all dollar amounts paid directly by the patient and his/her family as a result of cancer and cancer treatment, including wage losses but excluding out-of-pocket expenses directly related to treatment. For example, these costs include travel expenses related to surgery, nutrition expenses related to adjuvant therapies and time costs. All costs will be reported in 2014 Chinese yuan.

Specification of study secondary outcome variables

Delays in care. Currently, there are no accepted determinations of what constitutes a 'delay' regarding specific time intervals. However, a meta-analysis indicated that a delay of 3–6 months between the appearance of symptoms and the initiation of treatment among women with breast cancer was associated with a lower survival rate than a delay of <3 months.³⁷ Studies examining delay in breast cancer treatment have used the latter study as a template to define clinical delay.^{38–39} Delay of care will be modelled as continuous (number of days) and categorically: <1, 1 to 2 and 3+ months. For example, diagnosis and treatment delay of breast cancer will be the period between the date of the first physician visit related to breast cancer symptoms and a diagnosis date; and the interval between diagnosis date (month/year) and date of surgery.

Use of new screening and treatment methods. In addition to well-established guided care recommendations, the literature suggests patterns of new screening and treatment for study. For example, we will examine the early adoption of low-dose CT in lung cancer screening. The results of the National Lung Screening Trial published in 2011 demonstrated that screening for lung cancer using low-dose CT improved OS and reduced lung cancer mortality in the 55–74 years old age group by increasing the proportion of cancers detected at an early stage.⁴⁰ Other examples include examining use of combination chemotherapy within 60 days of surgery for hormone receptor-negative breast cancer <1 cm; axillary node dissection or sentinel node biopsy for stage I–IIb breast cancer; and adjuvant systemic therapy (combination chemotherapy and/or tamoxifen) for women older than 50 years with positive nodes, where chemotherapy will be identified from discharge and claims data.

Independent variables

Patient attributes and geographic location: The primary source of demographic data: for example, patient age, rurality of residence, sex and ethnicity will be the registry data and supplemented (health insurance) as needed with information from the claims and discharge data. Patient comorbidity will be examined as a predictor of care pattern by searching claims and discharge data for the diagnosis for various conditions during the year before the diagnosis of cancer, as suggested by Klabunde.⁴¹ We will also try to acquire the data of the number of oncologists and 2010 census population size by county to determine the ratio of oncologists to the population.

Location and distance to community oncology practices: We will identify oncology and radiation oncology services and locations. Cancer centres providing clinical care in Shandong province as of 2012 will be geocoded using standard geographic information system software. Geographic locations will be determined to calculate average travel time to area cancer care resources. Ground distances that reflect travel time to hospital facilities (based on topology) will be calculated.

Hospital characteristics: Public data provide institutional characteristics including grade/class, profit structure, occupancy, the volume of patients, diagnostic and therapeutic radiology services, chemotherapy services, outpatient services and other services, staffing, the number of beds, and the number of patients. In this database, each hospital is defined as either a teaching hospital or a community hospital based on their affiliation with medical schools.

Statistical analyses

There are four study populations: patients with lung, female breast, and gastrointestinal cancer and the non-cancer group in Shandong province. The primary outcome variable of the study is whether a patient received guideline-concordant care for lung, breast, and gastrointestinal cancer and the related costs. Secondary outcomes include timely care and use of newer treatments. For the primary outcomes, cancer incidence, stage at diagnosis, cancer care costs and the concordant rate of cancer care will first be described and documented for the study region, next hierarchical regression models will be constructed to explain variations in the primary outcomes by studying major independent variables that exhibit considerable diversity in the study region. We will use estimated cancer care outcomes with 95% CIs for description and documentation. Logistic regression will be used to assess the effect of these characteristics on dichotomous outcomes and for adjusting confounders. A generalised linear regression model with a γ distribution will be used to estimate these effects on cancer care costs. An innovation of our study is, in addition to conventional analyses, we will conduct robust analyses designed to address geospatial and instrumental variable (IV) effects. Taken together, the

series of planned analyses will allow us to make conclusions about correlates of variation in lung, breast and gastrointestinal cancer care, but enhanced with analysis of variation in care as a function of geographic clusters (geospatial analysis), and the extent that these patterns withstand adjustment for referral bias or selection (IV analysis).

We will describe and map geographic variation in patterns of care within Shandong province by location and place characteristics such as rurality/urbanicity and the accessibility of cancer care resources. We will estimate the incidence rates, the proportion of patients diagnosed at late stage, costs of cancer care, the overall rate of stage-based treatment concordance with cancer screening and treatment guidelines together with 95% CIs ($\pm 3\%$ error bound is expected based on our planned sample sizes). Geographic variation in study outcomes will be presented by calculating the rates for individual counties. Logistic regression will be used to model the study outcomes on location, and place characteristics.

We will identify patient, tumour, provider and facility factors that contribute to significant variation in cancer care including (1) age, sex, insurance and comorbidity status; and (2) travel time/distance from patients residence to specific types of providers (eg, high volume of cancer care, comprehensive treatment centre and oncologist practice). Logistic regression and generalised linear regression will be used to model the study outcomes on these patient-level variables. To account for the clustering of patients in the same geographic location or by the same care provider, random-effects or multilevel logistic models will be used.

We will model the effects of differential distance to specific types of providers (eg, high volume of cancer care, comprehensive cancer services) on cancer care outcomes as an IV that corrects for selective referral on treatment for early and locoregional stage disease. To specify the potential effects of provider type on guideline-concordant care, we will initially estimate logit models predicting cancer care outcomes as a function of the independent variables, plus an indicator that distinguishes patients treated at comprehensive cancer centres. A positive and statistically significant coefficient on these indicators would provide initial support for the hypothesis that patients receive better care from these types of providers. However, because our initial estimates relating to provider type could be biased by selective referral, we will re-estimate these models using an IV technique. The implementation of this technique requires the identification of an 'instrument' (or instruments) that satisfies two conditions: (1) the instrument must be associated with provider type (ie, whether patients are treated in comprehensive cancer centres or not); and (2) it must be reasonable to assume that the instrument affects concordance with guidelines only through the choice of provider type, with no direct effect on study outcomes and no other indirect

association with unmeasured confounders. In keeping with other studies in the literature that have investigated questions similar to ours,^{42 43} we propose to use differential travel distance/time by type of provider as our instrument. The interaction of differential travel time with a date of diagnosis in the winter months is another possible instrument, as winter travel conditions would increase effective travel times to more distant facilities. Having selected a suitable instrument, we will implement the IV procedure by estimating a supplementary equation that predicts choice of provider type as a function of differential distance (the instrument) and other relevant variables. We will compare the results of two different procedures for obtaining an IV estimate of the effect of provider type on study outcomes (eg, guideline concordance) in the primary equation of interest: (1) replacing actual provider type with the predicted probability of provider type from the supplementary equation (a two-stage procedure) or (2) estimating both equations as bivariate probit models in a full information maximum likelihood (one-stage) procedure.

ETHICS AND DISSEMINATION

Procedures to ensure compliance with ethical standards

Data collection will strictly adhere to the standards required by the committee—signatory to the Declaration of Helsinki 1964. Great care will be taken to ensure the confidentiality and anonymity of participants. For the quantitative component of the project, only disaggregated hospital discharge data will be obtained from the participating hospitals, and only necessary registry indicators and health insurance claim indicators will be collected. For the patient interviews, informed consents will be sought from patients with cancer before starting the interviews. Only disaggregated data will be used in the publications resulting from this project. All data will be securely stored and password protected, accessible only to the research team. In accordance with the Human Research Ethics Committee at Shandong University, all data will be preserved for 5 years following the completion of the project and then destroyed.

Dissemination of findings

The patterns of cancer screening and treatment and disparities have not been rigorously examined. In addition, cancer care costs remain unexamined in China. This project will depict the pattern of cancer screening, incidence and treatment in China and enhance our understanding of causes of disparities in cancer control. We will identify new and innovative scientific opportunities to improve cancer control and reduce inequities in communities experiencing an excess burden of cancer. We will also train students and young investigators to be part of the next generation of health service researchers. This project will create the state-of-the-art data that reflect the linkage of cancer registry file and health insurance claims. From the policy perspective, this study

works as a demonstration project to draw policymakers' attention to cancer prevention and control.

The project has several feedback mechanisms for disseminating the project results locally, regionally, nationally and internationally. They are:

1. The final report for the funding agency summarising the project results will be turned into policy briefs targeted at the relevant province-level and central government departments responsible for cancer prevention, control and treatment, health insurance administration, and other interested stakeholders. The final results of the briefs will be disseminated to the relevant stakeholders towards the second half of the final year of the project. For example, NY will disseminate the research findings to the ministers of the National Health and Family Planning Commission of China.
2. A 1-day Knowledge Transfer Workshop will be held in the fourth quartile of the final year of the project to bring together stakeholders from relevant government departments at the prefecture, province and national levels, hospitals, and CDCs. The discussion will use the final policy briefs circulated before the workshop in order to focus the discussion on the policy and practical implications of the research results and ways forward for the cancer care studies in China. Stakeholder feedbacks will be summarised and integrated into the final report to be submitted to the funding agency.
3. A more detailed version of the final report will also be made available to stakeholders locally and nationally at the conclusion of the project.
4. Peer-reviewed publications documenting the key findings of the project will be used to disseminate the project results to stakeholders within and outside of China. Open access journals will be targeted for manuscript submission as a way of reaching the widest possible audience locally and internationally who do not have easy access to university libraries.
5. Presentations in national, regional and international conferences on health equity or cancer care to publicise the research findings. The conferences will also be used as a forum for the research team to link up with academics and other stakeholders nationally and internationally to learn about the experiences of cancer care in other health systems and generate new ideas about ways of moving forward with the cancer care in China.

Author affiliations

¹Department of Public Health Sciences, University of Virginia, Charlottesville, Virginia, USA

²Shandong Provincial Cancer Hospital Affiliated to Shandong University, Shandong Academy of Medical Sciences, Jinan, China

³Center for Health Management and Policy (Key Laboratory of Health Economics and Policy, National Health and Family Planning Commission), Shandong University, Jinan, China

⁴School of Public Health and Management, Binzhou Medical University, Yantai, China

Acknowledgements The authors would like to thank all the local participants from Jinan and project counties for their efforts in launching and implementing this ongoing study.

Contributors NY, JW, JG, RA and XS collaboratively conceptualised and designed the study. JW, YC, JY, HW, JG and XS carried out data collection. NY, YC, JY, HW and XS analysed and interpreted the data. All authors have been involved in drafting the manuscript and have given final approval of the final version of this article.

Funding This study is funded by the China Medical Board (CMB 13-160).

Disclaimer The views expressed are those of the authors themselves and not necessarily the authors' institutions.

Competing interests None declared.

Patient consent Obtained.

Ethics approval The Institutional Review Board of the School of Public Health of Shandong University 20140201.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Data compiled from this project will be made available to all Chinese healthcare researchers.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

- Stewart BW, Wild CP. *World Cancer Report 2014*. Lyon, France: International Agency for Research on Cancer, 2014.
- He J, Gu D, Wu X, *et al*. Major causes of death among men and women in China. *N Engl J Med* 2005;353:1124–34.
- Hao J, Chen W. *Chinese cancer registry annual report*. Beijing: Military Medical Science Publishing House, 2012.
- Ramsey S, Blough D, Kirchoff A, *et al*. Washington State cancer patients found to be at greater risk for bankruptcy than people without a cancer diagnosis. *Health Aff (Millwood)* 2013;32:1143–52.
- Zhao P, Dai M, Chen W, *et al*. Cancer trends in China. *Jpn J Clin Oncol* 2010;40:281–5.
- Parkin DM, Bray F, Ferlay J, *et al*. Global cancer statistics, 2002. *CA Cancer J Clin* 2005;55:74–108.
- Calonge N, Petitti DB, DeWitt TG, *et al*. Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* 2009;151:716–26, W-236.
- Calonge N, Petitti DB, DeWitt TG, *et al*. Screening for colorectal cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* 2008;149:627–37.
- Moyer VA, Force USPST. Screening for cervical cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* 2012;156:880–91, W312.
- National Comprehensive Cancer Network. NCCN Guidelines™ for Breast Cancer. http://www.nccn.org/professionals/physician_gls/pdf/breast.pdf
- National Comprehensive Cancer Network. NCCN Guidelines™ for Non-Small Cell Lung Cancer. http://www.nccn.org/professionals/physician_gls/pdf/nscl.pdf
- National Comprehensive Cancer Network. NCCN Guidelines™ for Colon Cancer. http://www.nccn.org/professionals/physician_gls/pdf/colon.pdf
- National Cancer Institute. SEER-Medicare Publications by Journal & Year. http://healthservices.cancer.gov/seermedicare/overview/pubs_jour_year.php
- Jemal A, Center MM, DeSantis C, *et al*. Global patterns of cancer incidence and mortality rates and trends. *Cancer Epidemiol Biomarkers Prev* 2010;19:1893–907.
- Ferlay J, Shin HR, Bray F, *et al*. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* 2010;127:2893–917.
- Center MM, Jemal A, Ward E. International trends in colorectal cancer incidence rates. *Cancer Epidemiol Biomarkers Prev* 2009;18:1688–94.
- Jemal A, Bray F, Center MM, *et al*. Global cancer statistics. *CA Cancer J Clin* 2011;61:69–90.
- Zheng R, Zeng H, Zuo T, *et al*. Lung cancer incidence and mortality in China, 2011. *Thorac Cancer* 2016;7:94–9.
- Chen W, Zheng R, Baade PD, *et al*. Cancer statistics in China, 2015. *CA Cancer J Clin* 2016;66:115–32.
- Zheng R, Zeng H, Zhang S, *et al*. National estimates of cancer prevalence in China, 2011. *Cancer Lett* 2016;370:33–8.
- Song G, Meng F, Chen C, *et al*. A study of the relationship between cancer registration quality and population-based cohort. *China Cancer* 2009;18:531–2.
- Zhou H, Chi H, Peng J. Evaluation on cancer registry in Shenzhen, Guangdong province. *China Cancer* 2010;19:427–9.
- Huang Z, Chen W, Wu C, *et al*. Incidence and mortality of female breast cancer in China—a report from 32 Chinese cancer registries, 2003–2007. *Tumor* 2012;32:435–9.
- Hu S, Tang S, Liu Y, *et al*. Reform of how health care is paid for in China: challenges and opportunities. *Lancet* 2008;372:1846–53.
- The Xinhua News Agency. The basic health insurance plans covered 96% of the total population in China. http://www.gov.cn/jrzq/2012-07/27/content_2193635.htm
- Wagstaff A, Lindelow M, Jun G, *et al*. Extending health insurance to the rural population: an impact evaluation of China's new cooperative medical scheme. *J Health Econ* 2009;28:1–19.
- Sun X, Jackson S, Carmichael G, *et al*. Catastrophic medical payment and financial protection in rural China: evidence from the New Cooperative Medical Scheme in Shandong province. *Health Econ* 2009;18:103–19.
- Yu B, Meng Q, Collins C, *et al*. How does the New Cooperative Medical Scheme influence health service utilization? A study in two provinces in rural China. *BMC Health Serv Res* 2010;10:116.
- Andersen RM, McCutcheon A, Aday LA, *et al*. Exploring dimensions of access to medical care. *Health Serv Res* 1983;18:49–74.
- Albain KS, Green SR, Lichter AS, *et al*. Influence of patient characteristics, socioeconomic factors, geography, and systemic risk on the use of breast-sparing treatment in women enrolled in adjuvant breast cancer studies: an analysis of two intergroup trials. *J Clin Oncol* 1996;14:3009–17.
- Rosenblatt RA, Whitcomb ME, Cullen TJ, *et al*. Which medical schools produce rural physicians? *JAMA* 1992;268:1559–65.
- Kindig DA, Movassaghi H. The adequacy of physician supply in small rural counties. *Health Aff (Millwood)* 1989;8:63–76.
- Liu TG, Xiong SQ, Yan Y, *et al*. Use of Chinese herb medicine in cancer patients: a survey in south western China. *Evid Based Complement Alternat Med* 2012;2012:769042.
- Lou Y, Yates P, McCarthy A, *et al*. Fatigue self-management: a survey of Chinese cancer patients undergoing chemotherapy. *J Clin Nurs* 2013;22:1053–65.
- Short PF, Vasey JJ, Tunceli K. Employment pathways in a large cohort of adult cancer survivors. *Cancer* 2005;103:1292–301.
- Bradley CJ, Neumark D, Luo Z, *et al*. Employment and cancer: findings from a longitudinal study of breast and prostate cancer survivors. *Cancer Invest* 2007;25:47–54.
- Richards MA, Westcombe AM, Love SB, *et al*. Influence of delay on survival in patients with breast cancer: a systematic review. *Lancet* 1999;353:1119–26.
- Gwyn K, Bondy ML, Cohen DS, *et al*. Racial differences in diagnosis, treatment, and clinical delays in a population-based study of patients with newly diagnosed breast carcinoma. *Cancer* 2004;100:1595–604.
- Gorin SS, Heck JE, Cheng B, *et al*. Delays in breast cancer diagnosis and treatment by racial/ethnic group. *Arch Intern Med* 2006;166:2244–52.
- Aberle DR, Adams AM, Berg CD, *et al*. National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* 2011;365:395–409.
- Klabunde CN, Potosky AL, Legler JM, *et al*. Development of a comorbidity index using physician claims data. *J Clin Epidemiol* 2000;53:1258–67.
- Newhouse JP, McClellan M. Econometrics in outcomes research: the use of instrumental variables. *Annu Rev Public Health* 1998;19:17–34.
- Hadley J, Yabroff KR, Barrett MJ, *et al*. Comparative effectiveness of prostate cancer treatments: evaluating statistical adjustments for confounding in observational data. *J Natl Cancer Inst* 2010;102:1780–93.