BMJ Open

Contributions of prognostic factors to socioeconomic disparities in cancer survival: protocol for analysis of a cohort with linked data

Xue Qin Yu,[1,2] David Goldsbury,[1] Sarsha Yap,[1] Mei Ling Yap,[1,2,3,4] Dianne L O'Connell[1,2,5]

For numbered affiliations see end of article.

Correspondence to
Dr Xue Qin Yu;
xueqiny@nswcc.org.au

ABSTRACT

Introduction Socioeconomic disparities in cancer survival have been reported in many developed countries, including Australia. Although some international studies have investigated the determinants of these socioeconomic disparities, most previous Australian studies have been descriptive, as only limited relevant data are generally available. Here, we describe a protocol for a study to use data from a large-scale Australian cohort linked with several other health-related databases to investigate several groups of factors associated with socioeconomic disparities in cancer survival in New South Wales (NSW), Australia, and quantify their contributions to the survival disparities.

Methods and analysis The Sax Institute's 45 and Up Study participants completed a baseline questionnaire during 2006–2009. Those who were subsequently diagnosed with cancer of the colon, rectum, lung or female breast will be included. This study sample will be identified by linkage with NSW Cancer Registry data for 2006–2013, and their vital status will be determined by linking with cause of death records up to 31 December 2015. The study cohort will be divided into four groups based on each of the individual education level and an area-based socioeconomic measure. The treatment received will be obtained through linking with hospital records and Medicare and pharmaceutical claims data. Cox proportional hazards models will be fitted sequentially to estimate the percentage contributions to overall socioeconomic survival disparities of patient factors, tumour and diagnosis factors, and treatment variables.

Ethics and dissemination This research is covered by ethical approval from the NSW Population and Health Services Research Ethics Committee. Results of the study will be disseminated to different interest groups and organisations through scientific conferences, social media and peer-reviewed articles.

Strengths and limitations of this study

► The use of several linked health-related datasets will provide almost complete coverage of the care pathways for patients with cancer from prediagnosis to diagnosis and treatment and to the end of life for those who died.
► Information on lifestyle factors will be included in the analysis separately from other patients' characteristics.
► Both individual and neighbourhood socioeconomic measures will be used in the analysis, and their effects will be estimated independently and jointly.
► Residual confounding may still exist due to missing data for some measured factors and lack of data for patients' compliance with prescribed treatment, which may introduce bias through incomplete control of confounding.
► While the cohort may not be completely representative of the general New South Wales population (eg, the cohort is older and more educated), a recent study indicated that there was little evidence of bias in the association between the area-based socioeconomic measure and cancer survival in this cohort.

survival inequalities defined by either socioeconomic groups[5 7 10 11] or rural versus metropolitan residence.[9 11] Recent studies have also indicated that these inequalities in cancer survival have either persisted[12 13] or seem to be widening.[14 15] Although the underlying causes of these socioeconomic inequalities are not well understood, the possible reasons can be divided into three main groups[16]: factors related to tumour characteristics at diagnosis, factors related to treatment access and quality, and patients' characteristics. The latter group, including marital status, private health insurance coverage and comorbidities, may affect cancer outcomes through interacting with screening and treatment access and decisions. Identifying these factors and

INTRODUCTION

Even in developed countries with well-established healthcare systems, cancer survival is known to vary by socioeconomic level of patients.[1–4] In Australia, socioeconomic disparities in cancer survival have been reported over the past decade,[5–11] with

understanding how they may impact on cancer detection, treatment and survival are the crucial first steps towards addressing and removing these inequalities.

Several international studies have investigated the contributions of these determinants to survival disparities.[17–19] Stage at diagnosis and treatment were found to be important factors contributing to the survival differences between population groups for prostate[17] and breast cancer,[19] whereas these factors contributed only minimally to the association between socioeconomic position and colorectal cancer mortality.[18] These results suggest that the reasons for survival disparities are likely to vary with the country's population and health system characteristics, as well as by cancer type. Also, in Australia, some studies using linked datasets from multiple health-related data sources have attempted to disentangle the possible reasons for the poorer cancer outcomes for patients with lower socioeconomic backgrounds.[20 21] However, these Australian studies are limited due to a lack of relevant detailed data.[20 21] To our knowledge, no study has systematically considered the contributions of various prognostic factors, including individual lifestyle factors, on the disparity in cancer survival between socioeconomic groups in Australia.

The Sax Institute's 45 and Up Study is an ongoing large-scale Australian cohort study of healthy ageing of over 266 000 individuals aged 45 years and over residing in New South Wales (NSW), Australia.[22] Individuals joined the study by completing a postal questionnaire and giving consent for linkage of their personal information to routinely collected health datasets.[22] Linking the 45 and Up Study questionnaire data with several health-related databases enables the examination of possible reasons for socioeconomic disparities in cancer survival and quantification of the relative contributions of patients' characteristics and other prognostic factors. Information on many lifestyle factors was collected in the baseline questionnaire, such as smoking status, drinking habits, physical activity and body mass index (BMI). While these lifestyle factors have been shown to be associated with the risk of developing cancer,[23 24] an emerging body of evidence suggests that these lifestyle factors may also be associated with prognosis for patients with prostate[17] and colorectal cancer.[25 26] Thus, in this study, we intend to examine potential reasons for socioeconomic disparities in cancer-specific survival for those who were diagnosed with an incident cancer (colon, rectum, lung and female breast) in the 45 and Up Study cohort and to quantify the contributions of patients' characteristics, tumour-related factors and cancer treatments received, to disparities in cancer survival between socioeconomic population groups.

METHODS
Study sample
The 45 and Up Study participants joined the study in 2006–2009 by completing a baseline questionnaire, which collected information on a range of personal characteristics. The original aim of the study was to provide researchers with timely and reliable information on a wide range of exposures and outcomes, which are important public health issues for the ageing population.[22] The overall participation rate was 18%, and people 80+ years of age and residents of rural and remote areas were oversampled, and the study sample represents about 11% of the total NSW population aged 45 years or older. Prospective participants were randomly sampled from the Medicare enrolment database held by the Department of Human Services (formerly Medicare Australia), which provides near-complete coverage of the population. All cohort members were followed up for health-related events through linkage with several population-wide administrative health data collections (figure 1). For each data collection, we are using the most recent data available at the time. The proposed study sample will include those participants who were diagnosed with cancer after they joined the study, as identified through linkage to the population-based cancer registry. We will include patients diagnosed with cancer of the colon (C18), rectum (C19-20), female breast (C50) or lung (C34)[27] in this study. Participants with any prior record of cancer will be excluded, along with people first diagnosed at death.

Data sources
45 and Up Study baseline questionnaire
The 45 and Up Study is described in detail elsewhere.[22] For this proposed study, data will be taken from the 45 and Up Study baseline questionnaire (https://www.saxinstitute.org.au/our-work/45-up-study/), which collected information on key variables such as height, weight, smoking status, family history of disease and levels of physical activity as well as a range of sociodemographic information. Variables to be used in this analysis are shown in table 1.

NSW Cancer Registry (NSWCR)
The NSWCR contains, by statutory requirement, records of people diagnosed with cancer in NSW since 1972.[28] Data for cancers diagnosed in 1994–2013 will be used to identify prevalent and incident cancer cases in the study cohort. The data extracted for incident cases (diagnosed after recruitment) will include cancer type, cancer stage at diagnosis, date of diagnosis and histology, as well as demographic variables including age at diagnosis and sex. NSWCR data will also be used to exclude people who were diagnosed with cancer before they joined the 45 and Up Study (prevalent cases). Incident cases can be identified up to 7 years after recruitment for those who joined the study in 2006 and up to 4 years for those recruited in 2009.

NSW Admitted Patient Data Collection (APDC)
The APDC (July 2001 to June 2016) contains information on all inpatient separations, recorded as episodes of
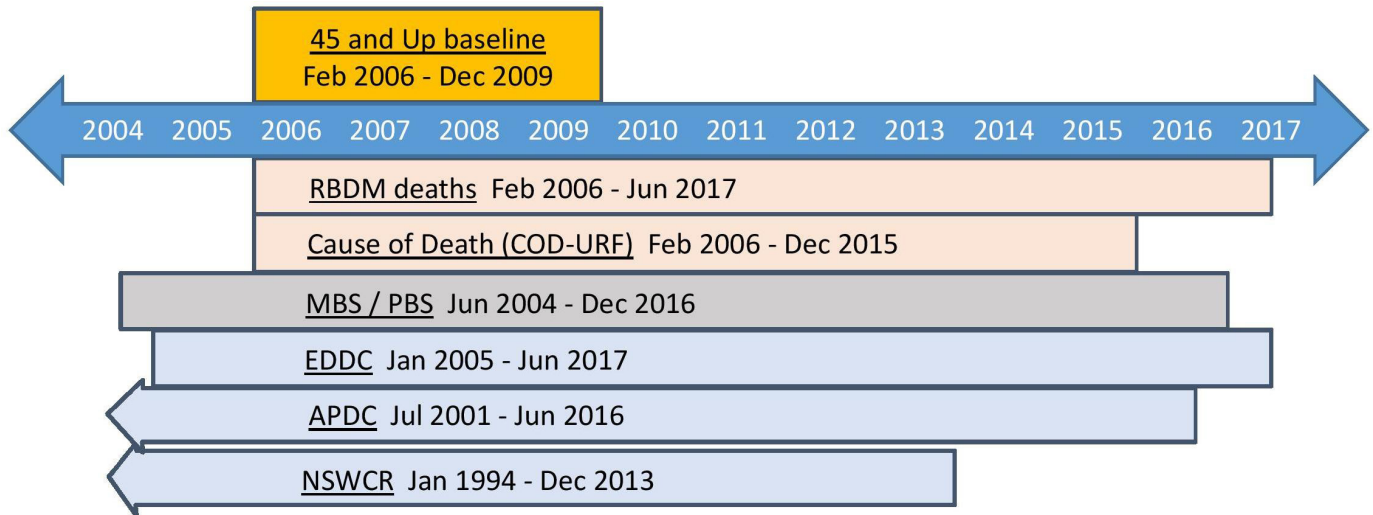
Figure 1 Health-related data collections and dates of availability. APDC, Admitted Patient Data Collection; COD-URF, Cause of Death Unit Record File; EDDC, Emergency Department Data Collection; MBS, Medicare Benefits Schedule; NSWCR, New South Wales Cancer Registry; PBS, Pharmaceutical Benefits Scheme; RBDM, New South Wales Registry of Births, Deaths and Marriages.

care, from all public, private and repatriation hospitals in NSW. It captures all procedures carried out and the diagnoses relating to the hospital episode of care. Patients with colon, rectal, breast and lung cancers are likely to use inpatient care, either for the first course of treatment, for complications with treatment or for other comorbid conditions. The APDC has been demonstrated to accurately provide information on surgery received by these patients.[29–31] Variables to be used in this analysis include dates of admission and separation, procedures carried

out and diagnoses relating to the hospital episode. APDC data will be used to capture surgery, systemic treatment and radiation therapy. In combination with Medicare Benefits Schedule (MBS) and Pharmaceutical Benefits Scheme (PBS) data, almost all episodes of these types of treatments will be captured.[31] In addition to identifying the cancer treatment received following diagnosis, these data will also be used to identify key chronic comorbid conditions recorded before their cancer diagnosis.

NSW Emergency Department Data Collection (EDDC)

The EDDC provides information on emergency department (ED) presentations to NSW public hospitals. This information will be used to identify whether the patient's diagnosis was preceded by an emergency presentation, which is associated with poorer survival even after taking into account cancer stage at diagnosis.[32] Variables to be used include dates of arrival and departure, ED visit type and mode of separation. The EDDC included data for over 80% of all ED presentations in NSW in 2006–2007, with near-complete coverage for metropolitan areas,[33] and by the end of the study period the EDDC had near-complete coverage of all public EDs in NSW, which are the vast majority of the hospital EDs in NSW.

MBS and PBS

The MBS is a database of Medicare services subsidised by the Australian government, which covers all Australian citizens and permanent residents. The PBS database is the administrative record of government subsidised medicines dispensed to all Australian citizens and permanent residents. The MBS and PBS data are both administered by the Department of Human Services. Radiation therapy (often conducted on an outpatient basis), systemic therapy and prescription medicines (chemotherapy and other cancer-related systemic therapies including
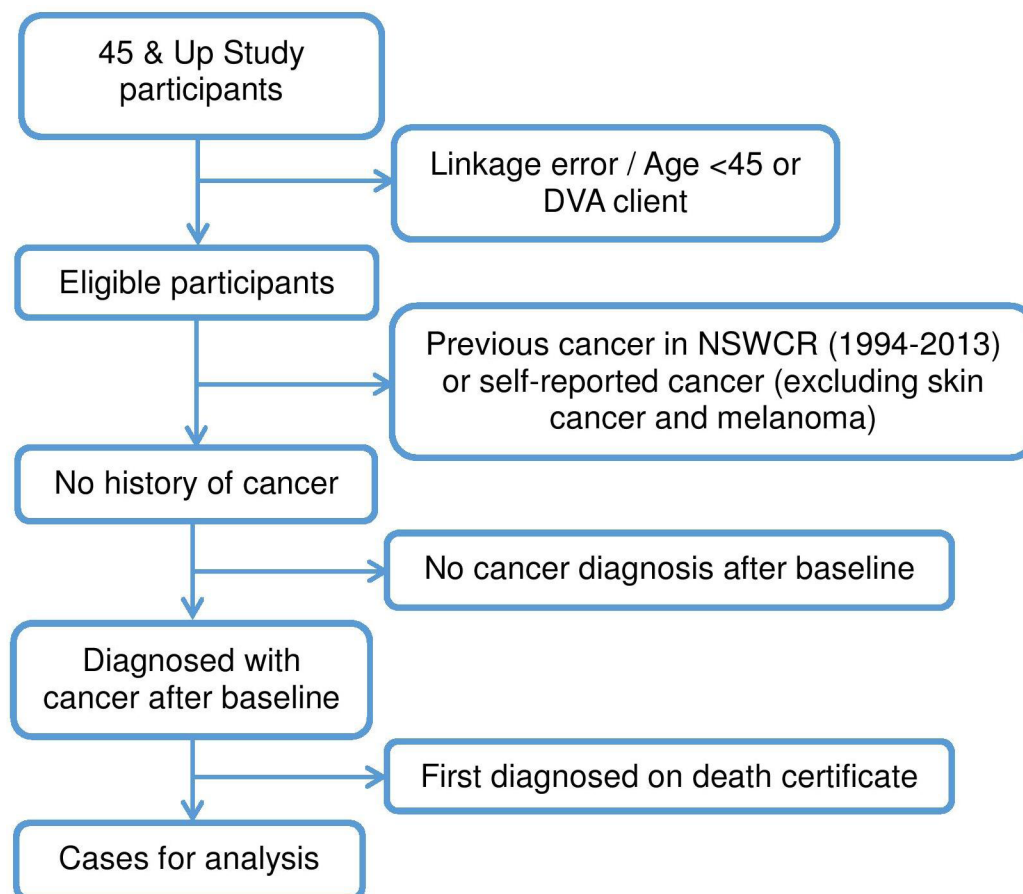
Table 1 Covariates to be included* in the analysis as prognostic factors for socioeconomic disparities in cancer survival

Patients' characteristics	Tumour-related/diagnosis-related factors	Treatment variables
a. Marital status	Cancer stage at diagnosis	Surgery
Private health insurance	Histology (varies by cancer type)	Systemic treatment
Place of residence	Emergency presentation	Radiation therapy
Comorbidity		
b. Tobacco smoking		
Alcohol consumption		
Physical activity		
Body mass index		

*These listed covariates will only be included in the analysis model if they are significantly associated with survival (with p<0.05) or they are important confounders of the socioeconomic position and survival association.

Figure 2 A flow chart of the study sample selection. DAV, Department of Veterans' Affairs; NSWCR, New South Wales Cancer Registry.

hormone therapy) can be identified through the MBS and PBS databases, respectively. Detailed names and PBS codes for medications relevant to these cancer types will be obtained from previous studies,[29–31] and in addition, we will identify and include new medications introduced for these cancers since the earlier publications. In addition, dates of service (MBS) and supply (PBS) will be used to measure the interval between cancer diagnosis and treatment receipt. Data available for analysis will be from June 2004 to December 2016 for participants joining the study prior to April 2008 and from September 2005 to December 2016 for all other participants. In this analysis, we cannot include participants whose healthcare was subsidised by the Australian Government's Department of Veterans' Affairs (identified by self-report or APDC/EDDC records), as their prescription medicines have a separate billing arrangement and these data are not available.

Death records from the NSW Registry of Births, Deaths and Marriages and the Australian Coordinating Registry's Cause of Death Unit Record File (COD-URF)

These death records contain information about date of death and cause of death for residents in NSW. This information will be used to determine length of survival for each individual patient, which is the main patient

outcome of interest in this study. Date of death information is available to June 2017 and cause of death data are available to December 2015 through linkage with the death records. COD-URF data include the underlying cause of death and up to 15 contributing causes of death.

A flow chart illustrating the selection of the study sample with inclusions/exclusions relevant to each data source is presented in figure 2.

Record linkage

The MBS and PBS data were linked to the 45 and Up Study cohort by the Sax Institute using a unique identifier that was provided to the Department of Human Services. Individual records from the NSWCR, APDC, EDDC and death datasets have been probabilistically linked to the 45 and Up Study cohort by the Centre for Health Record Linkage (CHeReL) using a best practice approach to linkage while preserving privacy.[34] A previous study found that the probabilistic linkage process was highly accurate with both false-positive and false-negative rates being <0.5%,[35] and for this linkage the CHeReL reported an estimated false-positive linkage rate of 0.5%.

Outcome variables

First, the cancer cases will be identified from the 45 and Up Study cohort by linkage with the NSWCR data. Then

their vital status and cause of death will be determined by linkage with the death data described above. Cancer-specific survival time will be estimated from the date of diagnosis to the date of death from the cancer under study or censored at the earlier date of death from another cause or at the end of follow-up (31 December 2015). To improve the accuracy of the estimates of cause-specific survival, we will use the Surveillance, Epidemiology, and End Results (SEER) cause-specific death classification, which classifies cancer deaths more accurately than the cause of death reported from death certificates.[36]

Study factors

The study factor of interest is patient's socioeconomic position measured in two ways. One measure is based on self-reported information provided in the baseline questionnaire (individual highest level of education), and the other is based on neighbourhood socioeconomic status (nSES). Each patient will be allocated to an nSES group based on the Statistical Area 1[37] they lived in at the time they joined the study. The nSES will be categorised into four groups based on quartiles of the state distribution of SES scores, according to the index of relative socioeconomic disadvantage from the 2011 Australian Census.[38] This index represents the average socioeconomic status of people living within a given neighbourhood in terms of level of material resources, relative to the whole population.[38] Highest education level will be categorised as: no school certificate, school certificate, higher school certificate or trade or certificate or diploma, and university degree or higher. To examine the combined effects of individual SES and nSES, a joint education and nSES variable will be created because there may be cross-level interaction between individual and neighbourhood SES on cancer survival.[39] For this variable, low education is defined as school certificate or below and low nSES as the two groups with lower score, resulting in four categories: 'high education and high nSES', 'high education and low nSES', 'low education and high nSES', 'low education and low nSES'.

Covariates

In our analysis, we will include multiple factors related to patients' characteristics, tumour-related factors and treatment variables, which were previously found to be associated with cancer survival.[17 25 26 40] The covariates to be included in this analysis are shown in table 1.

Factors related to patients' characteristics will be obtained from the 45 and Up Study baseline questionnaire, except for patients' comorbidities, which will be measured using the Charlson Comorbidity Index.[41] Using all available data in the APDC, comorbidities listed up to 5 years before the cancer diagnosis and up to 6 months after diagnosis will be included.[42] This index includes 17 medical conditions (excluding prior cancer and prior metastatic cancer) weighted with a score of 1–6 depending on the risk of dying associated with each one, so that each participant will have a total score calculated

based on the presence of these condition scores.[41] This index will be categorised into three groups for this analysis (0, 1 or ≥2). A sensitivity analysis will be performed by repeating the analysis using data on comorbidities up to 1 year prior to the cancer diagnosis.[43]

Marital status will be categorised as either married/de facto or other status. Private health insurance status will be defined as yes or no, with those having private hospital cover or having combined cover for private hospital and extras being classified as having private health insurance, as in a previous study.[44] Place of residence (based on Statistical Area 1 classification) at the time of enrolment for the study will be divided into major cities, inner regional and outer regional/remote/very remote using the Australian Standard Geographic Classification Remoteness Structure,[45] which classifies localities according to accessibility to major service centres based on road distances. Several lifestyle factors will also be included in the analysis. Tobacco smoking (current smoker, former smoker who quit in the past 15 years and never smoker or former smoker who quit >15 years ago), alcohol consumption (0, 1–14, >14 standard drinks per week) and physical activity will be defined based on self-reported data in the baseline questionnaire. Physical activity will be calculated using weighted minutes per week, adding up the number of minutes spent walking or being moderately active, and the number of minutes of vigorous activity multiplied by two, with the total being categorised as sedentary (0 min), insufficient (1–149 min), sufficient (150–299 min) or high (300+ minutes).[46] BMI will be calculated using self-reported height and weight (<18.5, 18.5–24.9, 25–29.9, ≥30 kg/m^2).[47]

Tumour-related or diagnosis-related factors will be derived from multiple data sources. Cancer stage at diagnosis as recorded in the NSWCR is grouped as localised, regional, distant or unknown based on pathology reports and statutory notifications to the registry.[12] Tumour histological types of interest will vary by cancer type. Whether the patient's diagnosis follows an emergency presentation (yes/no) will be determined by the date of diagnosis (from the Cancer Registry) and dates of ED arrival and departure (from the NSW EDDC). As there is no uniform definition of emergency presentation prior to cancer diagnosis,[32] we will repeat the analyses using alternative binary categories of time between emergency presentation and diagnosis (0–14 days vs 15 or more days, and 0–28 days vs 29 days or more, respectively).

Treatment information will also be derived from multiple data sources and be coded as yes/no based on any indication in any data source. First course of treatment will be defined as that commencing within 6 months of diagnosis, based on the date of diagnosis recorded in the Cancer Registry data. (We will perform a sensitivity analysis by repeating the analyses extending to 12 months after diagnosis to examine the effect of the timing of treatment received.) Receipt of radiation therapy (yes/no) will be determined using related information from the APDC and MBS data; receipt of systemic treatment

(yes/no) will be determined using related information from the PBS data, MBS data and APDC; receipt of cancer-directed surgery (yes/no) will be determined from the APDC (procedures carried out and diagnoses relating to the hospital episode) and MBS data (specific procedures). More details on how this treatment information is captured in the related data sources can be found in four previous publications.[29–31 48]

Patients will be classified as having surgical treatment if any procedure code International Statistical Classification of Diseases and Related Health Problems, Tenth Revision, Australian Modification (ICD10-AM) in any hospital separation indicated surgical removal of the cancer of interest or if the relevant surgical procedures were claimed through the MBS. The date of surgery will be taken as the date of admission for their first surgical procedure or the date of service recorded in the MBS data.

Patients will be classified as having cancer-related systemic treatment if (1) any procedure codes for any hospital separation indicated administration of cancer-related systemic therapy; (2) any diagnosis codes indicated admission for systemic therapy or (3) PBS records indicated systemic treatment was provided. The systemic therapy commencement date will be taken to be the first hospital admission date for systemic therapy or the first date of supply recorded in the PBS data, with hospital admission date taking precedence. The specific drugs of interest will be determined for the analysis for each cancer type, such as gemcitabine for lung cancer, fluorouracil for breast cancer and capecitabine for colon cancer.

Patients will be recorded as having received radiation therapy if: (1) it was indicated in the MBS dataset or (2) any procedure code in any hospital record indicated radiation therapy.[48] The radiation therapy commencement date will be taken to be the first date of service recorded in the MBS data or the date of hospital admission recorded in the APDC.

Statistical analysis

In this study, while more recent data for APDC, MBS/PBS and EDDC after 2015 are available, we will only use data up to 2015 as cause-specific survival is only available up to the end of 2015. For each cancer type, descriptive analyses of the distribution of factors among SES groups of interest will be undertaken using a χ^2 test. Then, Cox proportional hazards regression[49] will be used to examine associations between cancer-specific survival and socioeconomic position (with either highest nSES or those with a university degree or higher as the reference category). The baseline model will also include age at diagnosis (as a continuous variable), sex and year of diagnosis. The highest socioeconomic group will be used as the reference because that represents the desired survival rate for all groups to achieve.[50] Then, four more models will be fitted sequentially with one group of variables added at a time (groups shown in table 1). First, patient factors (demographic factors first, then lifestyle factors) will be added to the baseline

model (models 2a and 2b); then, tumour (cancer stage, histology) and diagnosis factors (emergency presentation) will be added to model 2b (model 3); finally, treatment variables will be added to model 3 (model 4). For each model, we will include nSES and education separately, then include them together in a single model, and finally we will model the joint effects of the two by including the combined variable. The significance of a single covariate or a group of covariates of interest when added to the previous model will be tested using a likelihood ratio test for the nested models. To identify the factors that by themselves have a significant influence on survival disparities, we will add these listed covariates individually to the baseline model (which includes age at diagnosis, year of diagnosis and sex). Covariates will only be included in the analysis model if they are significantly associated with survival (with p<0.05), and their addition does not change the estimated HRs for the socioeconomic position variables by more than 15%.

The contribution of each covariate will be calculated as in a previous study by Ellis et al.[40] First, total disparity (D_0) is defined as the HR for the socioeconomic group with the highest HR (usually the lowest socioeconomic group) versus the reference group, as derived from the baseline model including only SES, age at diagnosis, year of diagnosis and sex. Then, the change (usually decrease) in this disparity measure with the inclusion of each group of additional variables will be assessed by a measure of the relative change in disparity, based on the percentage of the total disparity that is explained by the addition of the new covariate(s) after accounting for the covariates that were already in the model. The measure is defined as $((D_- - D_+)/D_0) \times 100$, in which D_0 is the total disparity from the baseline model, D_- is the disparity measure from the model prior to introduction and D_+ is the disparity measure from the model after inclusion of the new covariate(s).

Patient and public involvement

Patient and public involvement in research is burgeoning in Australia, and we are still learning how best to involve consumers in research that is based on data analysis (and particularly in instances where we have not collected the data ourselves). Since developing this research protocol, we have reviewed our consumer involvement practices and implemented the following strategies to ensure the patient's perspective is considered in our future research: involve a dedicated, informed consumer in our research team to receive input into our research activities and request patients' feedback on questions we recommend should be incorporated into the 45 and Up Study.

Dissemination

Results of the study will be disseminated widely to different interest groups and organisations through scientific conference presentations, social media and peer-reviewed articles.

DISCUSSION

This study will provide new insights into the underlying reasons for survival differences between socioeconomic population subgroups in NSW, Australia. It is hoped that the findings from the study will be useful in suggesting and informing possible changes in health policy or service planning, which will ensure the best possible survival outcomes for all cancer patients, regardless of their socioeconomic background. Targeted interventions could potentially lead to a reduction (or elimination) of disparities in cancer survival between population groups, and thus, improved survival for the whole population. The proposed research aligns closely with goal 2 of the Cancer Institute NSW's state-wide plan for lessening the impact of cancers in NSW[51] and addresses one of the key areas relating to reducing survival disparities between population groups. Internationally, the proposed research will address the need to quantify the contribution of a range of factors to cancer survival disparities[16] and contribute to the literature on understanding the underlying reasons for such disparities.[17–19]

One important strength of the planned analyses is that we will systematically quantify the separate and combined contributions of multiple factors to socioeconomic disparities in cancer survival, allowing the identification of potential underlying reasons for the disparities so that appropriate interventions can be implemented to reduce these. Another strength of the study is that we can include lifestyle factors separately to other patients' characteristics such as marital status, private health insurance status and comorbidities, allowing us to investigate their independent impact on socioeconomic disparities in cancer survival. In addition, we will use both individual and neighbourhood socioeconomic measures in the analysis and estimate their effects both independently and jointly.

This study also has some potential limitations. These include the potential for residual confounding (not sufficiently adjusting for the measured factors due to missing data and measurement error), not being able to determine the appropriateness of care and having no information on patients' treatment choices or quality of life, or patients' compliance with prescribed treatments and clinical follow-up. In addition, the 45 and Up Study cohort is unlikely to be completely representative of the general population of NSW (eg, the cohort is older and more educated),[22] although a recent study indicated there was little evidence of bias in association between the area-based SES and cancer survival in this cohort.[52]

The methods we have developed for this planned analysis and the findings from the study will lead to a further programme of research including:

1. Developing appropriate intervention proposals to address any identified reasons for the disparities in survival between population groups.
2. Evaluating the effect of the interventions by monitoring outcomes specific to socioeconomic disparities.

3. Applying the methods developed to survival for other cancer types to obtain a broader understanding of the underlying reasons for survival disparities after a cancer diagnosis.

Author affiliations
¹Cancer Research Division, Cancer Council NSW, Sydney, New South Wales, Australia
²Sydney School of Public Health, University of Sydney, Sydney, New South Wales, Australia
³Ingham Institute for Applied Medical Research, University of New South Wales, Sydney, New South Wales, Australia
⁴Liverpool and Macarthur Cancer Therapy Centres, Western Sydney University, Campbelltown, New South Wales, Australia
⁵School of Medicine and Public Health, University of Newcastle, Newcastle, New South Wales, Australia

REFERENCES

1. Ellis L, Coleman MP, Rachet B. How many deaths would be avoidable if socioeconomic inequalities in cancer survival in England were eliminated? a national population-based study, 1996-2006. *Eur J Cancer* 2012;48:270–8.
2. Singh GK, Jemal A, Socioeconomic JA. Socioeconomic and racial/ethnic disparities in cancer mortality, incidence, and survival in the United States, 1950-2014: over six decades of changing patterns and widening inequalities. *J Environ Public Health* 2017;2017:2819372
3. Booth CM, Li G, Zhang-Salomons J, *et al*. The impact of socioeconomic status on stage of cancer at diagnosis and

survival: a population-based study in Ontario, Canada. *Cancer* 2010;116:4160–7.

4. Jansen L, Eberle A, Emrich K, *et al*. Socioeconomic deprivation and cancer survival in Germany: an ecological analysis in 200 districts in Germany. *Int J Cancer* 2014;134:2951–60.

5. Stanbury JF, Baade PD, Yu Y, *et al*. Impact of geographic area level on measuring socioeconomic disparities in cancer survival in New South Wales, Australia: a period analysis. *Cancer Epidemiol* 2016;43:56–62.

6. Yu XQ, O'connell DL, Gibberd RW, *et al*. A population-based study from New South Wales, Australia 1996-2001: area variation in survival from colorectal cancer. *Eur J Cancer* 2005;41:2715–21.

7. Yu XQ, O'Connell DL, Gibberd RW, *et al*. Assessing the impact of socio-economic status on cancer survival in New South Wales, Australia 1996-2001. *Cancer Causes Control* 2008;19:1383–90.

8. Yu XQ, O'Connell DL, Gibberd RW, *et al*. Estimating regional variation in cancer survival: a tool for improving cancer care. *Cancer Causes Control* 2004;15:611–8.

9. Jong KE, Smith DP, Yu XQ, *et al*. Remoteness of residence and survival from cancer in New South Wales. *Med J Aust* 2004;180:618–22.

10. Baade PD, Turrell G, Aitken JF. Geographic remoteness, area-level socio-economic disadvantage and advanced breast cancer: a cross-sectional, multilevel study. *J Epidemiol Community Health* 2011;65:1037–43.

11. Dasgupta P, Baade PD, Aitken JF, *et al*. Multilevel determinants of breast cancer survival: association with geographic remoteness and area-level socioeconomic disadvantage. *Breast Cancer Res Treat* 2012;132:701–10.

12. Stanbury JF, Baade PD, Yu Y, *et al*. Cancer survival in New South Wales, Australia: socioeconomic disparities remain despite overall improvements. *BMC Cancer* 2016;16:48.

13. Yu XQ, Luo Q, Smith DP, *et al*. Geographic variation in prostate cancer survival in New South Wales. *Med J Aust* 2014;200:586–90.

14. Yu XQ, Luo Q, Kahn C, *et al*. Temporal trends show improved breast cancer survival in Australia but widening urban-rural differences. *Breast* 2015;24:524–7.

15. Tervonen HE, Aranda S, Roder D, *et al*. Cancer survival disparities worsening by socio-economic disadvantage over the last 3 decades in New South Wales, Australia. *BMC Public Health* 2017;17:691.

16. Woods LM, Rachet B, Coleman MP. Origins of socio-economic inequalities in cancer survival: a review. *Ann Oncol* 2006;17:5–19.

17. DeRouen MC, Schupp CW, Koo J, *et al*. Impact of individual and neighborhood factors on disparities in prostate cancer survival. *Cancer Epidemiol* 2018;53:1–11.

18. Frederiksen BL, Osler M, Harling H, *et al*. Do patient characteristics, disease, or treatment explain social inequality in survival from colorectal cancer? *Soc Sci Med* 2009;69:1107–15.

19. Seneviratne S, Campbell I, Scott N, *et al*. Ethnic differences in breast cancer survival in New Zealand: contributions of differences in screening, treatment, tumor biology, demographics and comorbidities. *Cancer Causes Control* 2015;26:1813–24.

20. Beckmann KR, Bennett A, Young GP, *et al*. Sociodemographic disparities in survival from colorectal cancer in South Australia: a population-wide data linkage study. *BMC Health Serv Res* 2015;16:24.

21. Roder D, Zorbas HM, Kollias J, *et al*. Analysing risk factors for poorer breast cancer outcomes in residents of lower socioeconomic areas of Australia. *Aust. Health Review* 2014;38:134–41.

22. Banks E, Redman S, Jorm L, *et al*. Cohort profile: the 45 and up study. *Int J Epidemiol* 2008;37:941–7.

23. Cogliano VJ, Baan R, Straif K, *et al*. Preventable exposures associated with human cancers. *J Natl Cancer Inst* 2011;103:1827–39.

24. Stewart B, Wild CP. *World cancer report 2014*. Lyon: International Agency for Research on Cancer, 2014.

25. Boyle T, Fritschi L, Platell C, *et al*. Lifestyle factors associated with survival after colorectal cancer diagnosis. *Br J Cancer* 2013;109:814–22.

26. Lee J, Meyerhardt JA, Giovannucci E, *et al*. Association between body mass index and prognosis of colorectal cancer: a meta-analysis of prospective cohort studies. *PLoS One* 2015;10:e0120706.

27. World Health Organization. *Icd-10: international statistical classification of diseases and related health problems: tenth revision*. 2nd ed. Geneva: World Health Organization, 2004.

28. Cancer Institute NSW. *Nsw cancer registry*. Sydney: Cancer Institute NSW, 2018.

29. Goldsbury D, Weber M, Yap S, *et al*. Identifying incident colorectal and lung cancer cases in health service utilisation databases in Australia: a validation study. *BMC Med Inform Decis Mak* 2017;17:23.

30. Kemp A, Preen DB, Saunders C, *et al*. Ascertaining invasive breast cancer cases; the validity of administrative and self-reported data sources in Australia. *BMC Med Res Methodol* 2013;13:17.

31. Goldsbury DE, Armstrong K, Simonella L, *et al*. Using administrative health data to describe colorectal and lung cancer care in New South Wales, Australia: a validation study. *BMC Health Serv Res* 2012;12:387.

32. Zhou Y, Abel GA, Hamilton W, *et al*. Diagnosis of cancer as an emergency: a critical review of current evidence. *Nat Rev Clin Oncol* 2017;14:45–56.

33. O'Connell DL, Goldsbury DE, Davidson P, *et al*. Acute hospital-based services utilisation during the last year of life in New South Wales, Australia: methods for a population-based study. *BMJ Open* 2014;4:e004455.

34. SydneyCentre for Health Record Linkage, 2018NSW Ministry of Health. Available:

35. Bentley JP, Ford JB, Taylor LK, *et al*. Investigating linkage rates among probabilistically linked birth and hospitalization records. *BMC Med Res Methodol* 2012;12:149.

36. Howlader N, Ries LAG, Mariotto AB, *et al*. Improved estimates of cancer-specific survival rates from population-based data. *J Natl Cancer Inst* 2010;102:1584–98.

37. Australian Bureau of Statistics. *Australian statistical geography standard (ASGS*. Volume 1 ed. Canberra: ABS, 2016.

38. Australian Bureau of Statistics. *Information paper: an introduction to socio-economic indexes for areas (SEIFA), 2011*. Canberra: Australian Bureau of Statistics, 2013.

39. Chang C-M, Su Y-C, Lai N-S, *et al*. The combined effect of individual and neighborhood socioeconomic status on cancer survival rates. *PLoS One* 2012;7:e44325.

40. Ellis L, Canchola AJ, Spiegel D, *et al*. Racial and ethnic disparities in cancer survival: the contribution of tumor, sociodemographic, institutional, and neighborhood characteristics. *J Clin Oncol* 2018;36:25–33.

41. Charlson ME, Pompei P, Ales KL, *et al*. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987;40:373–83.

42. Yap S, Goldsbury D, Yap ML, *et al*. Patterns of care and emergency presentations for people with non-small cell lung cancer in New South Wales, Australia: a population-based study. *Lung Cancer* 2018;122:171–9.

43. Gibberd A, Supramaniam R, Dillon A, *et al*. Lung cancer treatment and mortality for Aboriginal people in New South Wales, Australia: results from a population-based record linkage study and medical record audit. *BMC Cancer* 2016;16:289.

44. Banks E, Jorm L, Lujic S, *et al*. Health, ageing and private health insurance: baseline results from the 45 and up study cohort. *Aust New Zealand Health Policy* 2009;6:17.

45. Australian Bureau of Statistics. *ASGC remoteness classification: purpose and use*. Canberra: Commonwealth of Australia, 2003.

46. Australian Government Department of Health. *Make your Move - Sit less - Be active for life!* Canberra: Australian Government Department of Health, 2014.

47. National Health and Medical Research Council. *Clinical practice guidelines for the management of overweight and obesity in adults, adolescents and children in Australia*. Melbourne: NHMRC, 2013.

48. Yap ML, O'Connell DL, Goldsbury D, *et al*. Comparison of four methods for estimating actual radiotherapy utilisation using the 45 and up study cohort in New South Wales, Australia. *Radiother Oncol* 2019;131:14–20.

49. Cox DR. Regression models and life-tables. *J R Stat Soc Ser B* 1972;34:187–202.

50. Harper S, Lynch J, Meersman SC, *et al*. An overview of methods for monitoring social disparities in cancer with an example using trends in lung cancer incidence by area-socioeconomic position and race-ethnicity, 1992-2004. *Am J Epidemiol* 2008;167:889–99.

51. Cancer Institute NSW. *Nsw cancer plan*. Sydney: Cancer Institute NSW, 2016.

52. Creighton N, Purdie S, Soeberg M, *et al*. Self-Selection in a population-based cohort study: impact on health service use and survival for bowel and lung cancer assessed using data linkage. *BMC Med Res Methodol* 2018;18:84.