Cancer Causes & Control
https://doi.org/10.1007/s10552-020-01361-6

ORIGINAL PAPER

# Effects of socioeconomic status on cancer patient survival: counterfactual event-based mediation analysis

**Shi-Woei Lin**[1,2] · **Kartika Nur Anisa**[1]

Received: 14 December 2019 / Accepted: 22 October 2020


## Abstract

**Purpose** This study investigated the direct and indirect effects of socioeconomic status (SES) on the survival time of cancer patients by using cancer stage to create a pathway from SES to health outcomes and facilitate a mechanistic inference.

**Methods** Both a traditional mediation analysis and a counterfactual event-based mediation analysis were applied to SEER (The Surveillance, Epidemiology, and End Results) data from the National Cancer Institute of the United States. A Cox proportional hazards model for survival analysis was performed in the mediation analysis.

**Results** The counterfactual event-based mediation analysis showed that the effect of SES on survival time was partially mediated by stage at diagnosis in lung (12%), liver (14.33%), and colorectal (9%) cancers. Investigation of the fundamental mechanism involved thus established the direct effect of SES on survival time and the indirect effect of SES on survival time through stage at diagnosis. Moreover, the mediation analysis also revealed that the disparity in timely diagnosis (i.e., stage at diagnosis) caused by SES was slightly significant.

**Conclusions** SES can either affect cancer survival directly or indirectly through stage at diagnosis. Opportunities to reduce cancer disparity exist in the design of early detection policies or mechanisms for patients with varying resources.

**Keywords** Cancer survival · Cox model · Mediation analysis · SEER

## Introduction

Approximately 9.5 million deaths in 2018 were attributable to cancer, of which lung cancer was the leading cause [1]. Other cancers with high mortality rates are liver, colorectal, stomach, breast, pancreas, esophagus, and prostate cancers [1]. Eradicating the burden of cancer and improving survival rates among cancer patients are therefore vital issues that need to be addressed by policy makers in the field of health care.

One critical factor that needs to be explored through cancer research is the length of time cancer patients can survive from the time of diagnosis to death, also known as survival time. The medical condition of a patient and the treatment they receive are among the factors that have a critical influence on survival time. However, socioeconomic conditions or the social economic status of a patient can also influence the prognosis of the disease. Over the last few years, several research studies have shown that poor socioeconomic status can have negative effects on cancer patient survival. For example, Kuhn et al. [2] found that both demographic factors and socioeconomic status (SES) influence the survival of pancreatic cancer patients. Other research has reported similar direct effects of SES on cancer survival (see, for example, [3–5]). Specifically, factors that contribute to disparities in survival for certain cancers (i.e., lung, female breast, prostate and colorectal cancers) include social position, social class, ethnicity, gender, and age. It should also be noted that while disparities in cancer survival may better be associated with variables related to individual SES, aggregated indicators or ecological-level data may help explain the role played by the social context, especially when the purpose of the study is to investigate the effects of the social procedure or impacts of health policies on survival [3].

SES may also have an indirect effect on cancer survival through multiple mediating variables. For example, low- and middle-income people may not have ready and easy access to health care facilities that would enable them to receive an

✉ Kartika Nur Anisa
d10701801@mail.ntust.edu.tw

[1] Department of Industrial Management, National Taiwan University of Science and Technology, Taipei, Taiwan

[2] Artificial Intelligence for Operations Management Research Center, National Taiwan University of Science and Technology, Taipei, Taiwan

Published online: 19 November 2020

early examination and/or better medication. Recent studies employing mediation analysis in this field have investigated the proportion of the effect that was indirect, namely to what extent a mediator variable was involved. For example, Lange and Hansen [6], Rochon et al. [7], and Akinyemiju et al. [8] have found that the effect of SES on survival time of cancer patients can be partially explained by mediator variables such as physical work environment [6], surgery and chemotherapy [7], and other cancer risk factors (e.g., nutrition and BMI) [8]. Among the mediator variables investigated, disease extension or cancer stage at diagnosis has been consider one of the most important prognostic determinant, and a considerable number of studies have investigated and confirmed the close relationship between SES, stage at diagnosis, and survival. Patients with lower SES usually link to a more advanced cancer stage at diagnosis in studies on lung, breast, colorectal, liver, prostate, stomach, and other cancers (see, for example, [9–13]). These studies underlined the relevance of stage as an important variable that related with socioeconomic class on survival. The indirect effects of SES on cancer survival (via different paths) pinpoint the essential mechanisms that require further investigation. This is because it is crucial for health care policy analysts to understand the causal mechanism that relates a proposed management intervention to a potential effect on patient survival so that they can ascertain the anticipated consequences of an action not yet taken.

To support management decisions regarding the allocation of health care resources, evaluating the direct or indirect effect of socioeconomic and demographic variables on survival time by using a decomposition approach can provide the high-quality information required.

The traditional approach to mediation analysis suggested by Baron and Kenny [14] involves estimating and comparing the effects of exposure on the outcome with and without adjusting for the mediator. However, applying the traditional approach to a model with nonlinear link functions between the exposure variables and the outcome variable may result in bias. Several researchers have identified the shortcomings of the approach suggested by Baron and Kenny [14]. However, in recent years, a mathematically consistent approach to mediation analysis based on a counterfactual framework has been developed. For example, a mediation analysis that can provide mathematically consistent interpretations of the mediation effect on a Cox regression survival model was developed by Lange et al. [15] by using the counterfactual event approach proposed by Pearl [16]. This effect is calculated by estimating the hazard ratios of covariate variables from a Cox proportional hazards model, which shows the proportion of the effect that can be explained by a mediator. This approach to mediation analysis is used when handling survival data because the Cox model can explain the multiplicative effect and important interactions between variables, even under different scales of effect.

This study therefore employed a counterfactual event-based mediation analysis to explore the effect of demographic and socioeconomic variables on the survival time of selected cancer cases (i.e., lung, liver, colorectal, and stomach cancers) through medical condition (i.e., stage at diagnosis). The main contribution of this study is enabling health care policy makers to allocate resources more effectively, reduce the burden of cancer, and improve cancer survival rates. This study also aims to demonstrate the usefulness of the counterfactual approach compared to the traditional mediation analysis using real data in medical field.

## Methods

### Data and variables

Data from The Surveillance, Epidemiology, and End Results (SEER) [17] project of the National Cancer Institute (NCI) of the United States were used for this investigation. This involved extracting data for specific cancer cases (i.e., lung, liver, colorectal, and stomach cancers) diagnosed between 1/2011 and 12/2014 and followed up for vital status (i.e., survival) until end of 12/2014. In total, 354,902 patient records were drawn from the SEER database and are shown in Table 1. The survival data of cancer patients are right-censored (i.e., patients might die or be lost to follow-up before the last date of study).

The medical condition (of cancer) is not only an essential factor that influences the prognosis of the disease but also a potential mediator that might be impacted by certain socioeconomic variables. This study used stage of cancer to denote medical condition, where each stage is classified into four levels: stage I, stage II, stage III, and stage IV. This is established from the size of the tumor and whether it has spread into nearby lymph nodes (or other parts of the body). Stage data records were defined by the American Joint Committee on Cancer (AJCC) used by SEER. Three demographic variables (age, gender, and race) were considered in this study. These demographic and socioeconomic variables were discretized, as shown in Table 1. Family income was divided into four categories based on the data quartile. Group 1 denoted the highest SES group and group 4 the lowest SES group. Family income data provided in the SEER database were area-level SES (for all the US counties and were collected from 2010 to 2014 Census Community Survey 5-year estimates). This indicator was linked to cancer cases or deaths at the county level, and the same value was used regardless of diagnosis/death years (since the study involved cancer cases diagnosed in the same time period). Thus, this study used the area-level data as the SES

**Table 1** Data Profile of Different Cancer Patients

| Characteristic | Lung Cancer | | | Liver Cancer | | | Colorectal Cancer | | | Stomach Cancer | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Number of patients | | % | Number of patients | | % | Number of patients | | % | Number of patients | | % |
| **Total Population** | | 183,867 | 100% | | 23,857 | 100% | | 127,038 | 100% | | 20,140 | 100% |
| **Year of Diagnosis** | | | | | | | | | | | | |
| 2011 | | 45,193 | 24.58% | | 5,470 | 22.93% | | 31,618 | 24.89% | | 4,854 | 24.10% |
| 2012 | | 46,151 | 25.10% | | 5,911 | 24.77% | | 31,891 | 25.10% | | 5,055 | 25.10% |
| 2013 | | 46,186 | 25.12% | | 6,178 | 25.90% | | 31,441 | 24.75% | | 5,090 | 25.27% |
| 2014 | | 46,337 | 25.20% | | 6,298 | 26.40% | | 32,088 | 25.26% | | 5,141 | 25.53% |
| **Age in year** | | | | | | | | | | | | |
| 34 | | 349 | 0.19% | | 187 | 0.78% | | 1,771 | 1.39% | | 340 | 1.69% |
| 35–44 | | 1,873 | 1.02% | | 401 | 1.68% | | 5,452 | 4.29% | | 904 | 4.49% |
| 45–54 | | 14,408 | 7.84% | | 3,231 | 13.55% | | 18,283 | 14.39% | | 2,453 | 12.18% |
| 55–64 | | 41,083 | 22.34% | | 9,687 | 40.60% | | 28,636 | 22.54% | | 4,498 | 22.33% |
| 65–74 | | 62,311 | 33.89% | | 5,966 | 25.01% | | 31,720 | 24.97% | | 5,443 | 27.03% |
| 75–84 | | 48,883 | 26.58% | | 3,476 | 14.57% | | 27,426 | 21.59% | | 4,605 | 22.86% |
| 85+ | | 14,960 | 8.14% | | 909 | 3.81% | | 13,750 | 10.83% | | 1,897 | 9.42% |
| **Gender** | | | | | | | | | | | | |
| Male | | 96,307 | 52.38% | | 18,041 | 75.62% | | 66,461 | 52.32% | | 12,949 | 64.29% |
| Female | | 87,560 | 47.62% | | 5,816 | 24.38% | | 60,577 | 47.68% | | 7,191 | 35.71% |
| **Race** | | | | | | | | | | | | |
| White/Others | | 162,957 | 88.63% | | 20,561 | 86.18% | | 111,968 | 88.14% | | 17,524 | 87.01% |
| Black | | 20,910 | 11.37% | | 3,296 | 13.82% | | 15,070 | 11.86% | | 2,616 | 12.99% |
| **Family Income** | | | | | | | | | | | | |
| 1 (highest) | ($27,120–$57,190) | 45,464 | 24.73% | ($27,120–$59,740) | 5,285 | 22.15% | ($27,120–$59,630) | 30,986 | 24.39% | ($27,120–$60,910) | 4,546 | 22.57% |
| 2 | ($57,191–$66,130) | 46,384 | 25.23% | ($59,741–$66,990) | 6,629 | 27.79% | ($59,631–$66,360) | 32,314 | 25.44% | ($60,911–$67,040) | 5,441 | 27.02% |
| 3 | ($66,131–$82,660) | 46,051 | 25.04% | ($66,991–$85,470) | 5,967 | 25.01% | ($66,361–$83,390) | 29,491 | 23.21% | ($67,041–$85,470) | 5,055 | 25.10% |
| 4 (lowest) | ($82,661–$126,040) | 45,968 | 25.00% | ($85,471–$126,040) | 5,976 | 25.05% | ($83,391–$126,040) | 34,247 | 26.96% | ($85,471–$126,040) | 5,098 | 25.31% |
| **Stage** | | | | | | | | | | | | |
| I | | 42,542 | 23.14% | | 9,674 | 40.55% | | 30,887 | 24.31% | | 6,013 | 29.86% |
| II | | 7,987 | 4.34% | | 4,610 | 19.32% | | 32,664 | 25.71% | | 2,690 | 13.36% |
| III | | 44,510 | 24.21% | | 5,358 | 22.46% | | 34,483 | 27.15% | | 2,194 | 10.89% |
| IV | | 88,828 | 48.31% | | 4,215 | 17.67% | | 29,004 | 22.83% | | 9,243 | 45.89% |
| **Survival Time in months** | | | | | | | | | | | | |
| (0–47) continuous | | 183,867 | 100% | | 23,857 | 100% | | 127,038 | 100% | | 20,140 | 100% |

indicator. Race was categorized into only two groups, Black and White/Others because the number of cases in "Other (American Indian/AK Native, Asian/Pacific Islander)" race group was very small.

## Statistical methods

This section briefly explains methods employed in this study: survival analysis, traditional mediation analysis, and counterfactual event-based mediation analysis.

(1) Survival Curve and Residual Life Time Estimation
Right-censored, time-to-event data were used in this study. The survival curve was estimated and graphed using the Kaplan–Meier (KM) method [18, 19].

(2) Mediation Analysis
The Cox proportional hazards regression model was used in both (i.e., traditional and counterfactual-based) mediation analyses for analyzing time-to-event data which record the length of time from diagnosis to the (possibly censored) disease endpoint. The Cox proportional hazards model is commonly used for dealing with survival data in medical literature and can explain multiplicative effects under the different effect scale in mediation analysis [20, 21]. Cox regression estimates the hazard ratios and the values were then used to determine the effect of the mediator variable between SES and the survival time of cancer patients.

### Traditional mediation analysis

The traditional approach to mediation analysis proposed by Baron and Kenny [14] can only be applied to the special case of linear models without interactions. In this approach, the first Cox model of the outcome variable $Y$ on the exposure variable $a$ and the confounding variable $c$ are formulated as (1).

$$E[Y|a,c] = \phi_0 + \phi_1 a + \phi_2 c. \quad (1)$$

The coefficient $\phi_1$ represents th total effect of exposure $a$ on outcome $Y$. The second regression model is then formulated by including mediator variable $m$, as shown in (2):

$$E[Y|a,m,c] = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 c. \quad (2)$$

If the exposure coefficient $\theta_1$ in (2) (i.e., with mediators added) decreases, compared with the exposure coefficient $\phi_1$ in (1), this confirms the existence of an indirect mediation effect because the mediator $m$ explains some of the effect of the exposure on the outcome $Y$ [15, 22]. Total effect (TE) is decomposed into direct and indirect effects. The exposure coefficient $\theta_1$ is defined as the direct effect (DE). The difference between $\phi_1$ and $\theta_1$ is an indirect effect (IE) or mediation effect (i.e., $IE = \phi_1 - \theta_1$).

### Counterfactual event-based mediation analysis

Lange et al. [15] proposed a new approach to mediation analysis based on the concept of a counterfactual event suggested by Pearl [16]. In this study, we investigated how the categorical mediator (i.e., stage at diagnosis) affected the relationship between SES and the survival time of cancer patients. Family income (one of the most important SES indices) was chosen as the focal exposure variable and grouped into four categories (with "1" being the highest). The mediator, cancer stage at diagnosis, was also grouped into four categories ("IV" being the most severe). Other independent or causal variables (i.e., age, gender, and race) that significantly affect survival time were also included in the model. Based on the proposed model in Lange et al. [15], the causal structure mediation analysis used in this study is presented in Fig. 1.

In a mediation analysis based on counterfactual framework, the direct effect is the effect observed on outcome variable by changing the exposure but keeping the mediator fixed at whatever level it would take if the exposure was kept on the original reference level. Similarly, for a specific exposure level, the indirect effect is the effect observed by changing the mediator as if we had changed the exposure (from the reference level to that specific level). Thus, for evaluating the indirect effect of SES on cancer survival through the mediator (i.e., stage at diagnosis) using counterfactual event approach, we need to first set the SES at a specific level $a$, and then evaluate the contrast effect between the counterfactual outcome if the mediator (i.e., stage at diagnosis) assumed whatever value it would have taken when the SES = $a$ and the counterfactual outcome if the mediator (i.e., stage at diagnosis) assumed whatever value it would have taken when the SES = $a^*$, where $a^*$ is the reference level. In other words, we need to simulate (via Monte Carlo

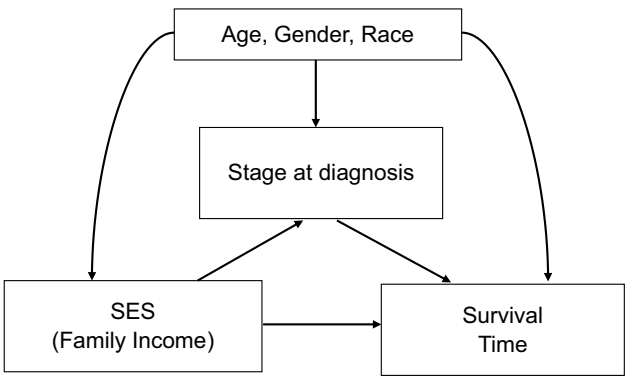


**Fig. 1** Directed acyclic graph of the causal structure mediation analysis

simulation) how a change in SES would affect the mediator for each individual, and how this change would in turn affect cancer survival.

The procedures involved in conducting the mediation analysis were as follows:

1. Using the original dataset with variable *A* representing the socioeconomic status (SES), estimate an appropriate model for the mediator (i.e., stage at diagnosis) conditional on SES indicator and the baseline confounders (i.e., age, gender, and race). Because the mediator is a categorical variable, a multinomial logistic regression model is constructed to estimate the effects.
2. Construct a new dataset by repeating each observation in the original dataset four times and include a new categorical variable *A** which capture four different possible values (of the categorical SES indicator) relative to the indirect path.
3. Compute the weight using the predicted functionality and the fitted model from step 1 on the data constructed in step 2. In this study, an approach of weighting the mediator probability distribution in counterfactual mediation analysis, proposed by Hong [23], is used to accommodate exposure (SES)–mediator (stage) interaction and adjust for the levels of SES (i.e., *A* vs. *A**) to estimate the marginal mean of each counterfactual outcome. In other words, for estimating the marginal mean, the weight is calculated as the ratio of the conditional probability of mediator (stage) under the new condition (new dataset, *A**) to the conditional probability of the mediator value under the original condition (SES indicator, *A*).
4. Fit a Cox model to the outcome involving *A, A*,* and other baseline confounders, and weight the model using the weights from the previous step to estimate DE and IE (from HRs of *A* and *A**, respectively). Then TE is the multiplicative from DE and IE.

## Results

This section presents and compares results obtained from two different statistical approaches to mediation analysis: the traditional approach (cox-based mediation analysis) or

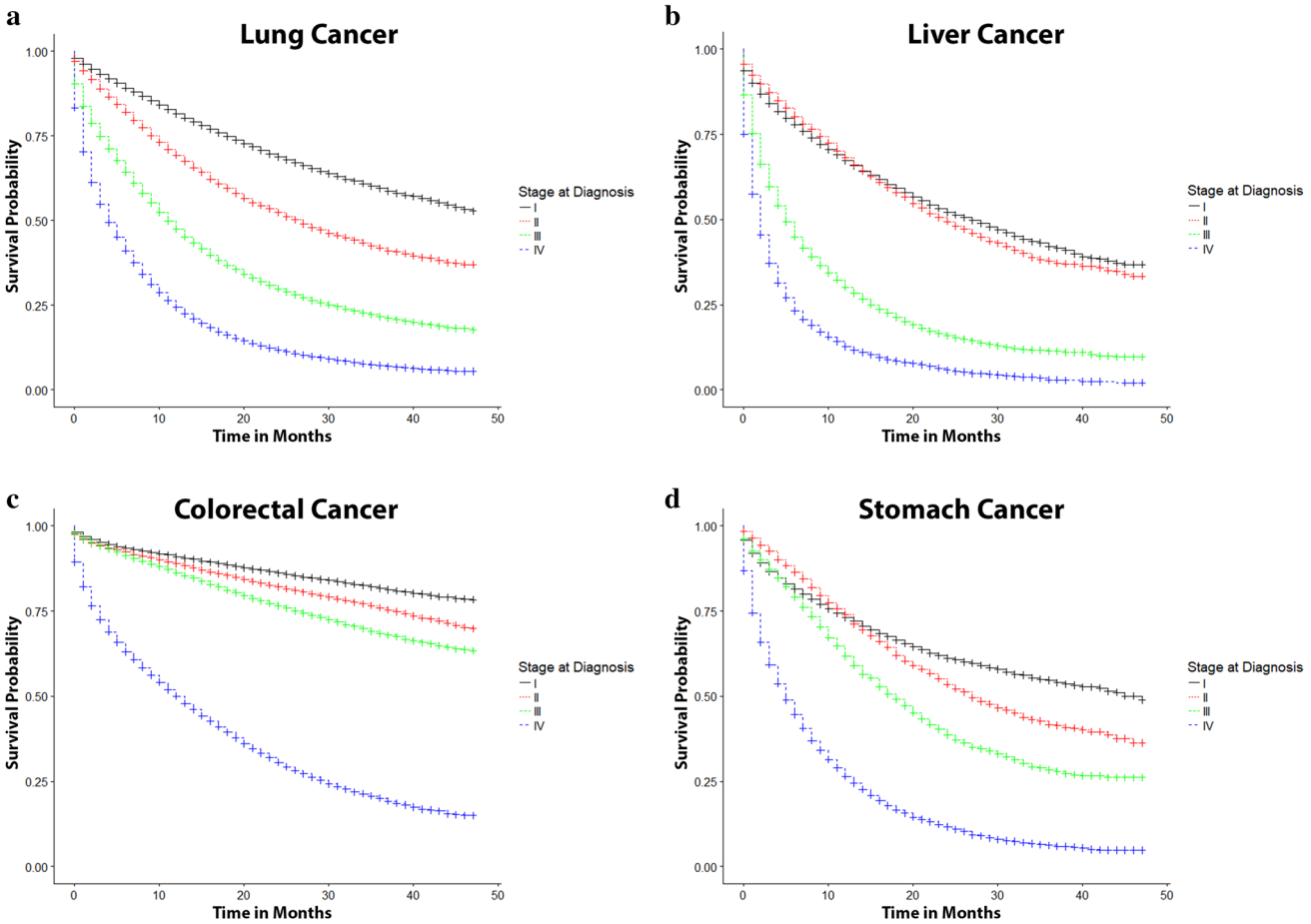


**Fig. 2** The survival curves based on stage at diagnosis of different cancer cases

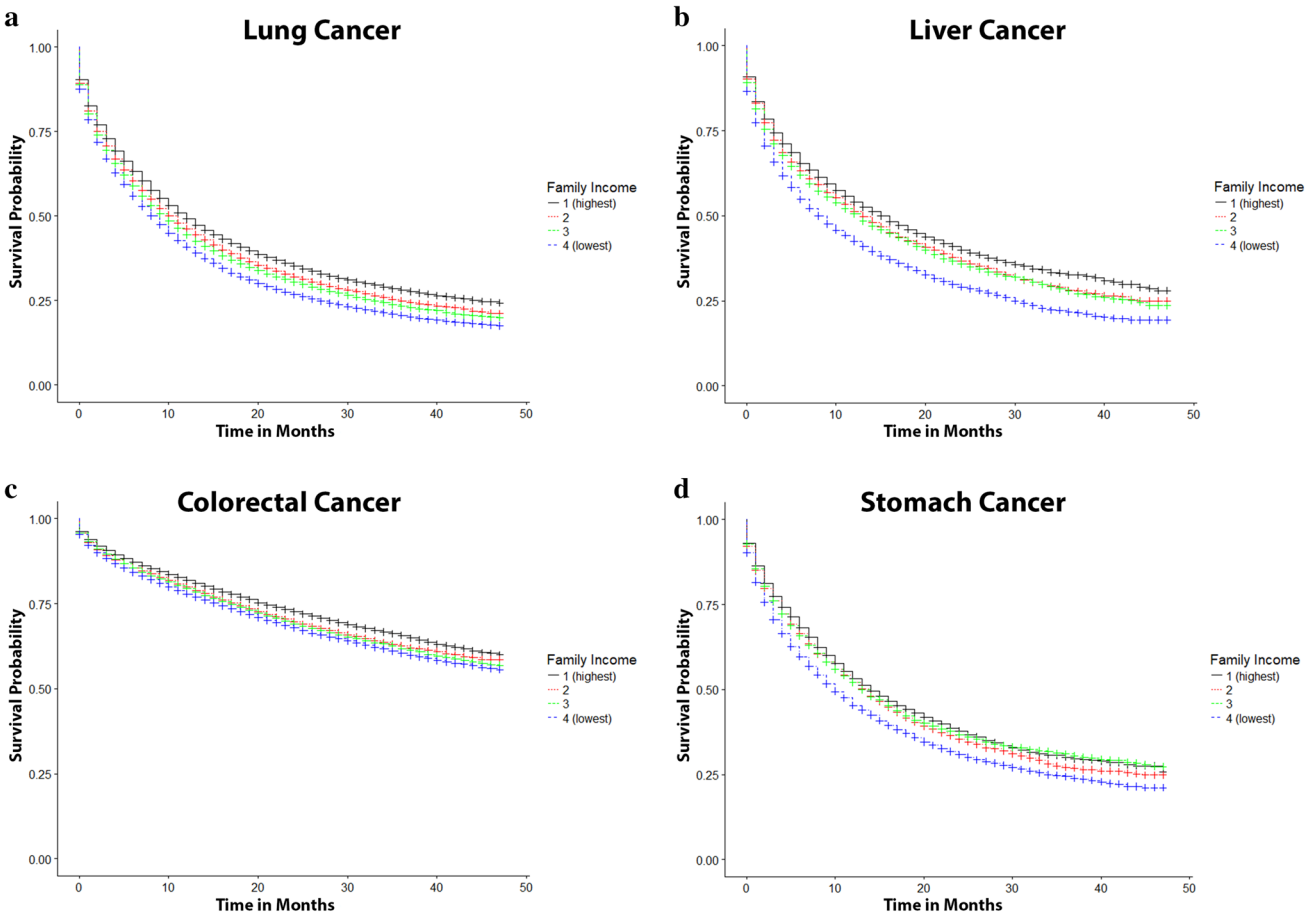


**Fig. 3** The survival curves based on socioeconomic status (i.e., family income) of different cancer cases

the counterfactual event-based approach. First, we graphically present the residual lifetime estimated using the Kaplan–Meier method to explore the relationship between the exposure SES variable, the mediator, and survival time. All analyses were performed using R statistical software.

## Residual lifetime estimation

To explore the effect of the SES level and cancer stage at diagnosis on survival time, survival curves were created for patients with differing SES levels and stages at diagnosis. These are presented in Figs. 2 and 3. The two-year survival rates for all types of cancer by stage and family

**Table 2** Two-year survival rate by stage at diagnosis and by socioeconomic status (i.e., family income)

| | Lung Cancer | Liver Cancer | Colorectal Cancer | Stomach Cancer |
|---|---|---|---|---|
| **Stage at Diagnosis** | | | | |
| I | 68.80% | 52.10% | 86.10% | 61.10% |
| II | 52.40% | 49.40% | 82.10% | 53.50% |
| III | 29.90% | 15.90% | 76.60% | 38.60% |
| IV | 11.80% | 5.80% | 30.60% | 11.60% |
| **Socioeconomic Status (Family Income)** | | | | |
| 1 (highest) | 35.20% | 39.90% | 72.70% | 37.70% |
| 2 | 32.00% | 36.70% | 69.70% | 35.40% |
| 3 | 30.50% | 35.80% | 69.10% | 36.80% |
| 4 (lowest) | 26.80% | 29.00% | 67.90% | 30.80% |

**Table 3** Hazard ratios (HRs) and 95% confidence intervals for cancer survival using cox models, by socioeconomic status (i.e., family income), unadjusted and adjusted by mediator (i.e., stage at diagnosis)

| SES Groups | HR | 95% CI | | SES Groups | HR | 95% CI | | Change | |
|---|---|---|---|---|---|---|---|---|---|
| **Lung Cancer** | | | | | | | | | |
| **Without Adjustment for Mediator** | | | | **With Adjustment for Mediator** | | | | | |
| 1 (highest) | 1.000 | | | 1 (highest) | 1.000 | | | | |
| 2 | 1.105 | 1.087 | 1.124 | 2 | 1.115 | 1.097 | 1.134 | 0.010 | 0.91% |
| 3 | 1.157 | 1.138 | 1.177 | 3 | 1.136 | 1.117 | 1.155 | 0.021 | 1.82% |
| 4 (lowest) | 1.288 | 1.266 | 1.309 | 4 (lowest) | 1.280 | 1.259 | 1.302 | 0.008 | 0.61% |
| Concordance= 0.579 | | | | Concordance= 0.713 | | | | | |
| **Liver Cancer** | | | | | | | | | |
| **Without Adjustment for Mediator** | | | | **With Adjustment for Mediator** | | | | | |
| 1 (highest) | 1.000 | | | 1 (highest) | 1.000 | | | | |
| 2 | 1.100 | 1.047 | 1.156 | 2 | 1.082 | 1.030 | 1.136 | 0.018 | 1.66% |
| 3 | 1.135 | 1.079 | 1.194 | 3 | 1.099 | 1.045 | 1.156 | 0.036 | 3.19% |
| 4 (lowest) | 1.387 | 1.320 | 1.457 | 4 (lowest) | 1.289 | 1.227 | 1.355 | 0.098 | 7.03% |
| Concordance= 0.569 | | | | Concordance= 0.709 | | | | | |
| **Colorectal Cancer** | | | | | | | | | |
| **Without Adjustment for Mediator** | | | | **With Adjustment for Mediator** | | | | | |
| 1 (highest) | 1.000 | | | 1 (highest) | 1.000 | | | | |
| 2 | 1.112 | 1.079 | 1.147 | 2 | 1.106 | 1.073 | 1.141 | 0.006 | 0.52% |
| 3 | 1.147 | 1.111 | 1.183 | 3 | 1.153 | 1.118 | 1.190 | 0.006 | 0.57% |
| 4 (lowest) | 1.233 | 1.196 | 1.270 | 4 (lowest) | 1.252 | 1.215 | 1.291 | 0.020 | 1.61% |
| Concordance= 0.64 | | | | Concordance= 0.781 | | | | | |
| **Stomach Cancer** | | | | | | | | | |
| **Without Adjustment for Mediator** | | | | **With Adjustment for Mediator** | | | | | |
| 1 (highest) | 1.000 | | | 1 (highest) | 1.000 | | | | |
| 2 | 1.080 | 1.024 | 1.139 | 2 | 1.081 | 1.024 | 1.140 | 0.001 | 0.08% |
| 3 | 1.049 | 0.993 | 1.108 | 3 | 1.086 | 1.028 | 1.147 | 0.037 | 3.49% |
| 4 (lowest) | 1.271 | 1.204 | 1.341 | 4 (lowest) | 1.325 | 1.255 | 1.399 | 0.055 | 4.29% |
| Concordance= 0.573 | | | | Concordance= 0.719 | | | | | |

income are presented in Table 2. Figure 2 shows that cancer patient survival differs markedly according to stage at diagnosis. The low level of stage (i.e., level I) has the highest survival probability than the other level stage at diagnosis. The survival rates of the stage level in all cancer cases are quite much different. Compared with the effect of stage at diagnosis on survival, survival rates for cancer patients with different levels of SES (i.e., family income) do not differ significantly (see Fig. 3 and Table 2). However, survival rates are still higher for patients with the highest level of SES (i.e., Level 1). This may be because patients with a high level of family income may be able to obtain more advanced treatments and therefore survive longer than patients with a lower level of family income.

## Traditional mediation analysis

The traditional mediation analysis shows that the hazard ratios (HRs) of lung and liver cancers are lowered after stage at diagnosis is included as the mediator variable, as shown in Table 3. For lung cancer, the HRs fall in SES group 3 and group 4. Although the HR of SES group 2 increases marginally after the mediator is added, the results still indicate that SES has a beneficial influence on patient survival (survival time). For liver cancer, HRs decrease for all SES groups. The changes indicate that the effect of SES on survival time is partially explained by the mediator, namely stage at diagnosis. For lung cancer, the average percentage change in the HR when family income is used as a SES indicator is 1.11%. This means that 1.11% of the effect of SES in lung cancer can be explained by stage. The average percentage changes in the other HRs are 3.96%, 0.9%, and 2.62% for liver, colorectal, and stomach cancers, respectively. The concordance statistic, a popular measure for evaluating goodness-of-fit in survival analysis, is also reported for each individual model in Table 3. The concordance statistic can be used to quantify the probability that the predicted survival time goes in the same direction as the actual data. With concordance statistics higher than 70%, substantial prognostic ability is validated for the Cox models (with adjustment for mediator).

Odds ratios (ORs) and 95% confidence intervals are generated to evaluate the associations between SES and stage

**Table 4** Odds ratios (ORs) and 95% confidence intervals for stage by socioeconomic status (i.e., family income)

| SES Groups | OR | 95% CI | | Significance |
|---|---|---|---|---|
| **Lung Cancer** | | | | |
| 1 (highest) | 1.000 | | | |
| 2 | 1.008 | 0.984 | 1.033 | 0.500 |
| 3 | 1.095 | 1.068 | 1.122 | 0.000 |
| 4 (lowest) | 1.092 | 1.066 | 1.119 | 0.000 |
| **Liver Cancer** | | | | |
| 1 (highest) | 1.000 | | | |
| 2 | 1.037 | 0.970 | 1.107 | 0.285 |
| 3 | 1.075 | 1.005 | 1.150 | 0.036 |
| 4 (lowest) | 1.180 | 1.103 | 1.263 | 0.000 |
| **Colorectal Cancer** | | | | |
| 1 (highest) | 1.000 | | | |
| 2 | 1.043 | 1.014 | 1.072 | 0.003 |
| 3 | 1.021 | 0.992 | 1.050 | 0.158 |
| 4 (lowest) | 1.018 | 0.990 | 1.047 | 0.209 |
| **Stomach Cancer** | | | | |
| 1 (highest) | 1.000 | | | |
| 2 | 1.027 | 0.954 | 1.105 | 0.481 |
| 3 | 0.976 | 0.906 | 1.052 | 0.533 |
| 4 (lowest) | 0.971 | 0.900 | 1.047 | 0.438 |

Note: The SES group is significant if $p\ value < 0.05$

**Table 5** Total effect (TE), direct effect (DE), indirect effect (IE) and proportion mediated (PM), with adjusted by mediator (i.e., stage at diagnosis)

| | TE | DE | IE | PM (log IE/log TE) |
|---|---|---|---|---|
| **Lung Cancer** | | | | |
| 1 → 2 | 1.104 | 1.101 | 1.003 | 0.03 |
| 1 → 3 | 1.156 | 1.121 | 1.031 | 0.21 |
| 1 → 4 | 1.287 | 1.250 | 1.030 | 0.12 |
| **Liver Cancer** | | | | |
| 1 → 2 | 1.099 | 1.088 | 1.011 | 0.11 |
| 1 → 3 | 1.133 | 1.109 | 1.022 | 0.17 |
| 1 → 4 | 1.390 | 1.322 | 1.052 | 0.15 |
| **Colorectal Cancer** | | | | |
| 1 → 2 | 1.112 | 1.093 | 1.018 | 0.17 |
| 1 → 3 | 1.146 | 1.136 | 1.009 | 0.06 |
| 1 → 4 | 1.232 | 1.223 | 1.008 | 0.04 |
| **Stomach Cancer** | | | | |
| 1 → 2 | 1.080 | 1.070 | 1.009 | |
| 1 → 3 | 1.049 | 1.058 | 0.992 | |
| 1 → 4 | 1.270 | 1.283 | 0.990 | |

at diagnosis. While there are 4 levels in stage, for better emphasize the effect of SES on cancer stage at diagnosis, we only provided the ORs for stage IV (the most critical level of stage) for different levels of SES. The significance of SES level in relation to stage is indicated by the ORs of stage by SES and the significance value of each SES group, as shown in Table 4. The ORs show that the lower the SES group, the higher the risk of being diagnosed at the latest stage, namely level IV. Table 4 shows that stage is significantly influenced by SES for lung cancer, but the difference between SES group 1 and 2 is not significant. The OR change from group 1 to group 2 increases the odds of being diagnosed at level IV of stage by 0.8% or 1.008 times. Similarly, in liver cancer, the ORs between SES group 1 and 2 are not significantly different, but stage is significantly influenced by SES because the odds of being diagnosed in the latest stage of cancer are significantly higher for SES groups 3 and 4 (when compared with group 1).

## Counterfactual event-based mediation analysis

The results for the counterfactual event-based mediation analysis, including the direct effects (DE) and indirect effects (IE) decomposed from total effects (TE), are shown in Table 5. As expected, high SES leads to a higher survival rate. For example, among lung cancer patients, a change from SES group 1 (the highest SES) to group 4 (the lowest SES) increases the hazard ratio of death by 28.7%. The changes from the highest SES to the lowest SES also significantly increase the HRs of death by 39%, 23.2% and 27% for liver, colorectal, and stomach cancers, respectively.

Furthermore, approximately 12% (the average of proportion mediated) of the effect of SES on lung cancer survival is mediated by stage, while the remaining 88% is explained by alternative pathways or a direct path. The mediating effect of changes in different levels of SES shows that the proportion of the effect explained by stage at diagnosis is considerably different.

The results also show that approximately 14.33% and 9% of the total effect of SES on liver and colorectal cancer survival, respectively, is mediated by stage. However, the remaining 85.67% (of liver cancer) and 91% (of colorectal cancer) are explained by alternative pathways or a direct path. The insignificant influence of SES to stage at diagnosis (see Table 4) and the opposite signs of DE and IE (see Table 5) for stomach cancer indicate that stage at diagnosis might not be the single mediator between SES and survival; therefore, the proportion mediated by cancer stage should not be computed.

## Discussion

From the results of residual lifetime estimation, survival rates for early-stage cancer are generally better than those for late-stage cancer. Thus, when patients are diagnosed as having the later stage (i.e., more severe) disease, they cannot survive longer than patients diagnosed as having the earlier stage disease. Otherwise, the highest level of SES (i.e., level 1) has the highest survival rates. This may be because patients with a high level of family income may be able to obtain more advanced treatments and therefore survive longer than patients with a lower level of family income.

From the results of the traditional mediation analysis, it is difficult to determine whether the effect of SES on survival time for colorectal and stomach cancers are partially influenced by stage at diagnosis. The results in Table 3 show that the HRs usually increase after the model is adjusted by the mediator (i.e., stage at diagnosis) for both cancers. Moreover, the ORs for stage by SES (see Table 4) for colorectal and stomach cancers are difficult to interpret because a higher level of family income was expected to give patients lower odds of being diagnosed at the most severe stage of cancer (i.e., level IV). Furthermore, this traditional approach generally underestimates the mediation effect and the results may be biased because the mediated effect cannot be shown clearly.

Mediation analysis based on the counterfactual event-based approach, however, produces better results. This approach clearly reveals how much of the effect of SES on survival time is directly contributed by the SES and how much is indirectly mediated by stage at diagnosis. For lung, liver, and colorectal cancers, the mediating effect can be easily interpreted as the results show that the effect of SES on survival time is partially mediated by stage, which is consistent with the literature [6–13]. A change in SES from high to low levels would increase the risk of death, and this finding is also consistently reported in literature [2–5]. However, the mediation analysis results for stomach cancer are rather difficult to interpret because of inconsistent mediation, and further investigation might still be needed. Some comparisons of the two mediation analysis approaches demonstrate that the traditional mediation analysis might not be an appropriate method with survival model and the corresponding time-to-event data used in this study because the traditional approach underestimates the mediation effects and the results lead to serious biases.

When patients have a high family income (or have lived in an area with higher family income), their chances of receiving an early diagnosis and/or obtaining more advanced medication are considerably higher. This finding seems to have been proved by some studies which found that the low-SES patients were less frequently referred to specialists or to oncology centers (i.e., branch of medicine which deals with the prevention, diagnosis, and treatment of cancer) [24–26]. Moreover, even with the same health care assessment given to the all SES level, low-SES patients were more likely diagnosed late-stage cancer at diagnosis. Some possible reasons have been stated by Singer et al. [27] that the patients were afraid of finding help or unable to explain the symptoms ("health illiteracy"). Another reason might be the physicians delayed referral to experts in those patients.

The mediation effects identified from this study suggest that policies that could make sure timely access of health care services or enable proactive screening and early diagnosis for deprived groups can still substantial improve cancer survival. Some studies focused mainly on part of the pathway of the mediation analysis conducted in the current study have also made the same conclusions. For example, they recommended the advancement made in cancer control should be given to the low-SES cancer patients, and improvements are clearly needed in cancer prevention, early detection, cancer screening, and treatment for the low-SES patient [27, 28]. Our mediation analysis investigated and confirmed the mechanism of the whole pathway. We thus believe that the policies might not only decrease cancer risk, but also might reduce the scale and costs of meditation considerably that would be far more beneficial, both economically and from a health care perspective.

Besides stage at diagnosis and SES, biologically factor (such as tumor biological features) might play an important role in explaining the cancer survival disparities [29]. Survival of patients after a cancer diagnosis is highly ruled by tumor biology [30]. However, stage at diagnosis is still the strongest prognostic determinant on cancer survival with SES disparities for solid tumors and frequently denotes delay in health-seeking or health supply by health care [31, 32].

## Conclusions

SES can affect cancer survival both directly and indirectly depending on the medical condition of the patient (i.e., stage at diagnosis). A counterfactual event-based mediation analysis provides clearer and more mathematically consistent results than the traditional approach and allows for better causal interpretation.

The results show there is a mediating effect in most cancer cases. The inconsistent level of mediation in stomach cancer makes it difficult to estimate the proportion mediated because the DE has the opposite sign to the IE (even though there is still a mediation effect in this case).

A limitation of this study lies in the fact that only area-level SES indices on family income are available for our investigation. While theoretically individual-leveled data

can better monitor the social inequalities, Quaglia et al. [3] pointed out that the ecological-level or aggregated indicators may still help to describe the effects of the socioeconomics context on cancer survival. Just as the study showed that the socioeconomic community can sometimes have significant impacts on individual health outcomes [33], the same could hold for area-level measurement of family income. The results of this study are consistent with the findings in previous studies that used area-level SES measures (e.g., household income [12, 34, 35] and family income [13, 36]) as indicators to evaluate the disparities in individual cancer outcomes. Furthermore, since this study clearly shows the mediating effect of SES on cancer survival via stage of diagnosis even when area-level SES indicator was used, it indicates that enforcing policies to provide health care resources for early detection of cancer can still significantly reduce the risk of cancer death. Even though area-level indices have been widely used to assess the effect of SES on cancer outcomes [3, 12, 13, 34–37], we still need to caution against over-interpretation of the effect. Individual-level SES data may still need to be considered in the future to better investigate social inequalities with regard to cancer risk.

Furthermore, since community SES indices may also not simply proxy for individual SES indices, it would be interesting to evaluate the role played by the context (or community) and the individual factors in the future. It is also possible to investigate the association of area-level SES and area-level health services supply on cancer outcomes in the future.

Another limitation of this study concerns the inconsistent mediation that was found. This problem may be resolved by adding other important mediators (e.g., treatment modalities or treatment procedures) and/or using multiple mediators, because opposing mediation relationships will be more easily observed when estimating multiple mediator models. Additionally, adding some variables (i.e., psychosocial and demographic variables such as educational, marital status, and others) as covariates might facilitate identification of a more effective causal mechanism related to SES and survival time.

**Funding** Funding was provided by Ministry of Science and Technology, Taiwan (MOST 103-2410-H-011-012-MY3, MOST 106-2410-H-011-004-MY3).

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

1. IARC (International Agency for Research on Cancer, World Health Organization) (2018) All cancers. http://gco.iarc.fr/today/data/factsheets/cancers/39-All-cancers-fact-sheet.pdf. Accessed 15 Nov 2018
2. Kuhn Y, Koscielny A, Glowka T, Hirner A, Kalff JC, Standop J (2010) Postresection survival outcomes of pancreatic cancer according to demographic factors and socio-economic status. Eur J Surg Oncol 36(5):496–500
3. Quaglia A, Lillini R, Mamo C, Ivaldi E, Vercelli M (2013) Socio-economic inequalities: a review of methodological issues and the relationships with cancer survival. Crit Rev Oncol Hematol 85(3):266–277
4. O'Keefe EB, Meltzer JP, Bethea TN (2015) Health disparities and cancer: racial disparities in cancer mortality in the United States, 2000–2010. Front Public Health 3:51
5. Robert SA, Strombom I, Trentham-Dietz A, Hampton JM, McElroy JA, Newcomb PA et al (2004) Socioeconomic risk factors for breast cancer: distinguishing individual- and community-level effects. Epidemiology 15(4):442–450
6. Lange T, Hansen JV (2011) Direct and indirect effects in a survival context. Epidemiology 22(4):575–581
7. Rochon J, du Bois A, Lange T (2014) Mediation analysis of the relationship between institutional research activity and patient survival. BMC Med Res Methodol 14(1):9
8. Akinyemiju T, Moore JX, Pisu M (2018) Mediating effects of cancer risk factors on the association between race and cancer incidence: analysis of the NIH-AARP Diet and Health Study. Ann Epidemiol 1(28):33–40
9. Kendra LS, Heather C, Fawn DV, Karl B, Mousumi B (2003) Race, socioeconomic status and stage at diagnosis for five common malignancies. Cancer Causes Control 14:761–766
10. Sun-Seog K, Min-Gyeong K, Mi-Ran K, Min-Ho S, Jin-Su C (2017) Difference of stage at cancer diagnosis by socioeconomic status for four target cancers of the National Cancer Screening Program in Korea: results from the Gwangju and Jeonnam cancer registries. Journal of Epidemiology 27:299–304
11. Timothy JC, Francis PB (2015) Urban/rural disparities in cancer incidence in New York State, 2008–2012. PeerJ PrePrints:e1500
12. Carlos ARO, Jean LF, Yong-Fang K, James SG (2007) The influence of marital status on stage at diagnosis and survival of older persons with melanoma. J Gerontol Ser A Biol Med Sci 62(8):892–898
13. Yabroff KR, Leon G (2003) Does stage at diagnosis influence the observed relationship between socioeconomic status and breast cancer incidence, case-fatality, and mortality? Soc Sci Med 57(12):2265–2279
14. Baron RM, Kenny DA (1986) The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol 51(6):1173–1182
15. Lange T, Vansteelandt S, Bekaert M (2012) A simple unified approach for estimating natural direct and indirect effects. Am J Epidemiol 176(3):190–195
16. Pearl J (2009) Causality: models, reasoning and inference. Cambridge University Press, New York
17. Surveillance Research Program, National Cancer Institute SEER*Stat software version 8.3.5. https://www.seer.cancer.gov/seerstat
18. Guyot P, Ades A, Ouwens MJ, Welton NJ (2012) Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan-Meier survival curves. BMC Med Res Methodol 12(1):9
19. Collett D (2003) Modelling survival data in medical research, 2nd edn. Chapman & Hall/CRC, Boca Raton

20. Vander Weele TJ (2016) Mediation analysis: a practitioner's guide. Annu Rev Public Health 37:17–32
21. Kartsonaki C (2016) Survival analysis. Diagn Histopathol 22(7):263–270
22. O'Quigley J (2008) Proportional hazards regression. Springer, New York
23. Hong G (2010) Ratio of mediator probability weighting for estimating natural direct and indirect effects. In: Proceedings of the American Statistical Association, biometrics section. Alexandria, American Statistical Association, pp 2401–2415
24. Herbert C, Lefèvre H, Gignoux M, Launoy G (2002) Influence of social and occupational class and area of residence on management and survival in patients with digestive tract cancer: a population study in the Calvados area (France). Rev Epidemiol Sante Publique 50(3):253–264
25. Blais S, Dejardin O, Boutreux S, Launoy G (2006) Social determinants of access to reference care centres for patients with colorectal cancer—a multilevel analysis. Eur J Cancer 42(17):3041–3048
26. Kingsmore D, Ssemwogerere A, Hole D, Gillis C (2003) Specialisation and breast cancer survival in the screening era. Br J Cancer 88(11):1708–1712
27. Singer S, Roick J, Briest S, Stark S, Gockel I, Boehm A et al (2016) Impact of socio-economic position on cancer stage at presentation: findings from a large hospital-based study in Germany. Int J Cancer 139(8):1696–1702
28. Bradley CJ, Given CW, Roberts C (2001) Disparities in cancer diagnosis and survival. Cancer 91(1):178–188
29. Yan B, Noone AM, Yee C, Banerjee M, Schwartz K, Simon MS (2009) Racial differences in colorectal cancer survival in the detroit metropolitan area. Cancer 115(16):3791–3800
30. Gore JL, Kwan L, Saigal CS, Litwin MS (2005) Marriage and mortality in bladder carcinoma. Cancer 104(6):1188–1194
31. Coleman MP, Gatta G, Verdecchia A, Esteve J, Sant M, Storm H et al (2003) EUROCARE-3 summary: cancer survival in Europe at the end of the 20th century. Ann Oncol 14(suppl_5):v128–v149
32. Woods LM, Rachet B, Coleman MP (2006) Origins of socioeconomic inequalities in cancer survival: a review. Ann Oncol 17(1):5–19
33. Robert SA (1999) Socioeconomic position and health: the independent contribution of community socioeconomic context. Annu Rev Sociol 25(1):489–516
34. Ripping TM, van der Waal D, Verbeek AL, Broeders MJ (2016) The relative effect of mammographic screening on breast cancer mortality by socioeconomic status. Medicine 95(31):e4335
35. Rodday AM, Parsons SK, Snyder F, Simon MA, Llanos AA, Warren-Mears V et al (2015) Impact of patient navigation in eliminating economic disparities in cancer care. Cancer 121(22):4025–4034
36. Robert SA, Strombom I, Trentham-Dietz A, Hampton JM, McElroy JA, Newcomb PA et al (2004) Socioeconomic risk factors for breast cancer: distinguishing individual-and community-level effects. Epidemiology 15(4):442–450
37. Clegg LX, Reichman ME, Miller BA, Hankey BF, Singh GK, Lin YD et al (2009) Impact of socioeconomic status on cancer incidence and stage at diagnosis: selected findings from the surveillance, epidemiology, and end results: National Longitudinal Mortality Study. Cancer Causes Control 20(4):417–435