

# Estimating and testing the influence of early diagnosis on cancer survival via point effects of diagnoses and treatments

Statistical Methods in Medical Research

2022, Vol. 31(8) 1538–1548

© The Author(s) 2022



Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/09622802221098429

[journals.sagepub.com/home/smm](https://journals.sagepub.com/home/smm)**Xiaoqin Wang<sup>1</sup>, Johannes Blom<sup>2</sup>, Weimin Ye<sup>3</sup>, and Li Yin<sup>3</sup>** 

## Abstract

A cancer diagnosis is part of a complex stochastic process, which involves patient's characteristics, diagnosing methods, an initial assessment of cancer progression, treatments and a certain outcome of interest. To evaluate the performance of diagnoses, one needs not only a consistent estimation of the causal effect under a specified regime of diagnoses and treatments but also reliable confidence interval, *P*-value and hypothesis testing of the causal effect. In this article, we identify causal effects under various regimes of diagnoses and treatments by the point effects of diagnoses and treatments and thus are able to estimate and test these causal effects by estimating and testing point effects in the familiar framework of single-point causal inference. Specifically, using data from a Swedish prognosis study of stomach cancer, we estimate and test the causal effects on cancer survival under various regimes of diagnosing and treating hospitals including the optimal regime. We also estimate and test the modification of the causal effect by age. With its simple setting, one can readily extend the example to a large variety of settings in the area of cancer diagnosis: different personal characteristics such as family history, different diagnosing procedures such as multistage screening, and different cancer outcomes such as cancer progression.

## Keywords

Blip effect, cancer diagnosis, causal effect, G-formula, point effect, treatment regime

## 1 Introduction

A cancer diagnosis is part of a complex stochastic process, in which patients' personal and social characteristics influence the choice of diagnosing methods, diagnosing methods in turn influence the initial assessment of cancer stage, cancer stage, in turn, influences the choice of treating methods and treating methods in turn influence cancer outcomes such as cancer survival. To evaluate the performance of cancer diagnoses, one needs to estimate and test the causal effect under a specified regime of diagnoses and treatments, shorthanded as sequential causal effect (SCE). An important example of the SCE is the one under the optimal regime.

In the causal inference of treatment sequence, Robins derived the well-known G-formula, which identifies the SCE under a regime of treatments via standard parameters.<sup>1,2</sup> Based on Robins' G-formula, a parametric likelihood-based approach has been developed for estimating the SCE via standard parameters.<sup>3,4</sup> If the time-dependent covariates between treatments are posttreatment variables from the previous treatments as well as confounders for the subsequent treatments, however, this approach may suffer from the curse of dimensionality and the null paradox.<sup>3,4</sup> To avoid these problems, semi-parametric non-likelihood-based approaches are developed, for instance, the marginal structural model

<sup>1</sup>Department of Electrical Engineering, Mathematics, and Science, University of Gävle, Sweden

<sup>2</sup>Department of Clinical Science and Education, Södersjukhuset, Karolinska Institutet, Stockholm, Sweden

<sup>3</sup>Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Sweden

## Corresponding author:

Li Yin, Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Sweden.

Email: [li.yin@ki.se](mailto:li.yin@ki.se)

– including the doubly robust estimation – based on the inverse probability of treatment weighting<sup>4–6</sup> and the G-estimation based on the structural nested mean model (SNMM) or optimal-regime SNMM.<sup>2,4,7</sup> In a recent review on the optimal regime, Kosorok and Laber<sup>8</sup> highlighted the need for reliable methods of evaluating the uncertainty in estimating SCEs (Section 6, statistical inference).

Recently, Wang and Yin<sup>9,10</sup> derived a new version of the G-formula, which identifies the SCE by the point effects of individual treatments in the sequence. The point effect of treatment is simply the one in the causal inference of single-point treatment. Its estimation is well studied in the framework of single-point causal inference, for instance, it can be likelihood-based and further doubly robust to possible misspecification of the outcome model or the treatment assignment model.<sup>11</sup> The estimation does not suffer from the curse of dimensionality and the null paradox. Based on the new G-formula, they proposed a parametric likelihood-based approach, in which they were able to estimate and test the SCE without suffering from the curse of dimensionality and the null paradox by estimating and testing the point effects. They achieved not only an unbiased estimate of the SCE but also the nominal level of coverage probability for the confidence interval.

In this article, we study the application of the parametric likelihood-based approach based on the new G-formula to the area of an early cancer diagnosis. We estimate and test SCEs under various regimes, including the optimal one, of diagnoses and treatments as well as the modifications of SCEs by covariates. We will illustrate our method by an interesting example using data from a Swedish prognosis study of stomach cancer and provide practical advice.

## 2 Data and the assumption for estimating and testing SCEs

In Sweden, patients usually seek medical help at hospitals near their residential areas (namely, catchment areas). When cancer is diagnosed, they may stay at the diagnosing hospital or transfer to another hospital for treatment. The hospital diagnosing cancer is called the diagnosing hospital, while the one treating cancer is called the treating hospital. To evaluate the performance of diagnosing and treating hospitals, we may study cancer outcomes under various regimes of diagnosing and treating hospitals among cancer patients after adjusting for patients' differences. A question of relevance to public health policy is which type of the diagnosing and treating hospitals, large versus small, performs better on cancer survival.

The data used in this study is from a prognosis study conducted during the period between 1988 and 1995 in hospitals in central and northern Sweden.<sup>12</sup> It contained information on 910 patients with stomach cancer. The large type refers to the regional or county hospitals and the small type to local hospitals. The diagnosing hospital is the treatment variable  $Z_1$ , which takes the value  $z_1 = 0$  for small type and  $z_1 = 1$  for large type. The treating hospital is the treatment variable  $Z_2$ , which takes  $z_2 = 0$  for small type and  $z_2 = 1$  for large type.

The following stationary covariates before  $Z_1$  are measured: gender  $X_{11}$ , geographic area  $X_{12}$  and age  $X_{13}$ . Gender is  $x_{11} = 0$  for females and  $x_{11} = 1$  for males. The geographic area is categorized into  $x_{12} = 0$  for rural and  $x_{12} = 1$  for urban. Age takes continuous values  $x_{13}$ . The time-dependent covariate between  $Z_1$  and  $Z_2$  is cancer stage  $X_2$  taking the values  $x_2 = 1, 2, 3, 4$ . The outcome of interest is  $Y$ , which takes  $y = 0$  for death and  $y = 1$  for survival within one year after diagnosis or within 3 years. The data and code for all analyses are available in Supplemental Data of this article. The descriptive statistics are given in Table 1.

Due to a long-term social welfare system and relatively uniform culture in Sweden, we assume that gender  $x_{11}$ , geographic area  $x_{12}$  and age  $x_{13}$  are the only confounders for  $Z_1$  and additional cancer stage  $x_2$  for  $Z_2$ . Together with positivity and consistency assumptions, this assumption of no hidden confounders is the identifying condition.<sup>2,3</sup> Under the identifying condition, we can estimate and test SCEs under various regimes of diagnosing and treating hospitals by using the observed data.

Because the cancer stage has a significant influence on diagnosing hospital as well as on treating hospital, one cannot use the standard methods such as the usual regression to estimate and test the SCE.<sup>1,2</sup> In the following, we will apply the new G-formula to estimate and test the SCE via the point effects of diagnosing and treating hospitals.

## 3 Estimating and testing SCE under the regime of diagnosing and treating hospitals on one-year survival

Wang and Yin<sup>9</sup> constructed point parametrization for the likelihood of all the observable variables arising from a sequence of treatments, with the point effects of treatments in the sequence as a subset of the parametrization. They showed that with this parametrization, the score functions for the point effects of treatments are little associated with one another at different treatment times, so that the point effects can be estimated separately at different treatment times, like the point effect of one single-point treatment. Furthermore, Wang and Yin<sup>10</sup> derived a new version of the G-formula, which expresses the point effects in terms of the blip effects of treatments and then all other SCEs in terms of the blip effects. As a result, a likelihood

**Table 1.** Frequencies or means (standard deviations) of covariates and outcome across the diagnosing and treating hospitals for 910 stomach cancer patients.

	Stomach cancer: frequency or mean (standard deviation) on levels of $z_1$ and $z_2$			
	$z_1 = 0$	$z_1 = 1$	$z_2 = 0$	$z_2 = 1$
$N$	371	539	327	583
$x_{11} = 0$	153	212	137	228
= 1	218	327	190	355
$x_{12} = 0$	215	259	184	290
= 1	156	280	143	293
$x_{13}$	71.38 (11.10)	71.76 (11.10)	71.96 (10.80)	71.41 (11.28)
$x_2 = 1$	56	100	49	107
= 2	43	55	41	57
= 3	111	117	93	135
= 4	161	267	144	284
One year survival				
$y = 0$	209	310	186	333
= 1	162	229	141	250
Three-year survival				
$y = 0$	292	413	259	446
= 1	79	126	68	137

- Hospital types:  $z_1 = 0, 1$  for small and large diagnosing hospitals;  $z_2 = 0, 1$  for small and large treating hospitals.
- Stationary covariates:  $x_{11} = 0, 1$  for female and male;  $x_{12} = 0, 1$  for rural and urban residential areas;  $x_{13}$  for age.
- Time-dependent covariate:  $x_2 = 1, 2, 3, 4$  for cancer stages.
- Outcome:  $y = 0, 1$  death and survival within one year after diagnosis or within 3 years.

approach is developed to estimating and testing the blip effect and the SCE via these point effects.<sup>9,10</sup> Here, we introduce a concrete procedure of applying their method to an observational study, where the treatment assignment mechanism is unknown and the covariates can be continuous.

### 3.1 Point effects of diagnosing and treating hospitals

Let  $\mu(x_{11}, x_{12}, x_{13}, z_1) = E(Y|x_{11}, x_{12}, x_{13}, z_1)$  be the mean of observable survival  $Y$  in the stratum  $(x_{11}, x_{12}, x_{13}, z_1)$ . In the causal inference of single-point treatment, the point effect of diagnosing hospital  $z_1$  in a stratum  $(x_{11}, x_{12}, x_{13})$  is

$$\vartheta(x_{11}, x_{12}, x_{13}; z_1) = \mu(x_{11}, x_{12}, x_{13}, z_1) - \mu(x_{11}, x_{12}, x_{13}, 0)$$

This point effect is a mixed effect of  $z_1$  and  $z_2$ , but it will be used to obtain SCEs below. Let  $\mu(x_{11}, x_{12}, x_{13}, z_1, x_2, z_2) = E(Y|x_{11}, x_{12}, x_{13}, z_1, x_2, z_2)$  be the mean of observed survival  $Y$  in the stratum  $(x_{11}, x_{12}, x_{13}, z_1, x_2, z_2)$ . Then the point effect of treating hospital  $z_2$  in a stratum  $(x_{11}, x_{12}, x_{13}, z_1, x_2)$  is

$$\vartheta(x_{11}, x_{12}, x_{13}, z_1, x_2; z_2) = \mu(x_{11}, x_{12}, x_{13}, z_1, x_2, z_2) - \mu(x_{11}, x_{12}, x_{13}, z_1, x_2, 0)$$

For a small sample, it would be efficient to use the logistic model and the Bernoulli-based distribution to estimate the point effect for a dichotomous outcome. For a large sample like ours, one may also use the linear model and the normal distribution for the estimation. For the simplicity of explication, we use linear model and normal distribution to estimate the point effect as follows.

To estimate the point effect of diagnosing hospital  $z_1$ , we model the mean  $\mu(x_{11}, x_{12}, x_{13}, z_1)$ , in which diagnosing hospital  $z_1$  is the variable of interest while all others are possible confounders of  $z_1$ . Notably, the model does not contain any posttreatment variable after  $z_1$ , so the null paradox does not necessarily occur and the model can be unsaturated without biasing the point effect. At the significance level of 0.10, the residential area  $x_{12}$  is not significant, in line of the medical observation that the residential area is not as influential as gender and age. So we remove it from the model. Furthermore, the interactions between  $z_1$  and  $(x_{11}, x_{12}, x_{13})$  are not significant either. Finally, we obtain

$$\mu(x_{11}, x_{13}, z_1) = \beta_1 + \beta_2 x_{11} + \beta_3 x_{13} + \vartheta_1 z_1 \tag{1}$$

From this model, we obtain the estimate for the point effect of  $z_1 = 1$ , that is,  $\vartheta(x_{11}, x_{13}; 1) = \mu(x_{11}, x_{13}, 1) - \mu(x_{11}, x_{13}, 0) = \vartheta_1$ , which is the same for all  $(x_{11}, x_{13})$ . The estimate  $\hat{\vartheta}_1$  is presented in Table 2.

**Table 2.** Point effects, blip effects and optimal SCEs of diagnosing and treating hospitals on one-year survival of stomach cancer: estimate, P-value and 95% CI.

	Estimate, P-value and 95% CI in column for various causal effects				
Point effect	$\vartheta_1$	$\vartheta_{21}$	$\vartheta_{22}$	$\vartheta_{23}$	$\vartheta_{24}$
	−0.007 0.819 (−0.071, 0.056)	0.005 0.918 (−0.083, 0.092)	0.089 0.357 (−0.100, 0.277)	0.027 0.682 (−0.103, 0.158)	−0.034 0.382 (−0.109, 0.042)
Blip effect	$\gamma_1$	$\gamma_{21}$	$\gamma_{22}$	$\gamma_{23}$	$\gamma_{24}$
	−0.009 0.710 (−0.054, 0.037)	0.005 0.918 (−0.083, 0.092)	0.089 0.357 (−0.100, 0.277)	0.027 0.682 (−0.103, 0.158)	−0.034 0.382 (−0.109, 0.042)
Optimal SCE	$\Gamma_1$	$\Gamma_{21}$	$\Gamma_{22}$	$\Gamma_{23}$	$\gamma_{24}$
	0.028 0.237 (−0.019, 0.075)	0.053 0.220 (−0.032, 0.138)	0.089 0.357 (−0.100, 0.277)	0.027 0.682 (−0.103, 0.158)	−0.034 0.382 (−0.109, 0.042)

Estimate, P-value and 95% CI for effect modification of large diagnosing hospital by age:  $\gamma_{1,2}$ : −0.001, 0.298, (−0.003, 0.001); for intercept:  $\gamma_{1,1}$ : 0.064, 0.378, (−0.078, 0.205).

Estimate, P-value and 95% CI for effect modification of large treating hospital by age:  $\gamma_{21,2}$ : 0.012, 0.037, (0.001, 0.023); for intercept:  $\gamma_{21,1}$ : −0.801, 0.030, (−1.526, −0.077).

- Five point effects:  $\vartheta_1$  of large versus small diagnosing hospital;  $\vartheta_{2j}$  of large versus small treating hospital for cancer stage  $j = 1, 2, 3, 4$ .
- Five blip effects:  $\gamma_1$  of large versus small diagnosing hospital;  $\gamma_{2j}$  of large versus small treating hospital for cancer stage  $j = 1, 2, 3, 4$ .
- Five optimal SCEs:  $\Gamma_1$  under optimal regime of diagnosing and treating hospital;  $\Gamma_{2j}$  of optimal treating hospital for cancer stage  $j = 1, 2, 3$ ;  $\Gamma_{24} = 0$  of optimal treating hospital for cancer stage 4 is the baseline, so the blip effect  $\gamma_{24}$  is instead presented.
- Modification of the blip effect by age:  $\gamma_{1,2}$  modifying the blip effect of large diagnosing hospital by age;  $\gamma_{1,1}$  the intercept of the blip effect;  $\gamma_{21,2}$  modifying the blip effect of large treating hospital for cancer stage 1 by age;  $\gamma_{21,1}$  the intercept for the blip effect.

To estimate the point effect of treating hospital  $z_2$ , we model the mean  $\mu(x_{11}, x_{12}, x_{13}, z_1, x_2, z_2)$ , in which treating hospital  $z_2$  is the variable of interest while all others are possible confounders of  $z_2$ . Notably, the model does not contain any posttreatment variable after  $z_2$ , so the null paradox does not necessarily occur and the model can be unsaturated without biasing the point effect. At the significance level of 0.10, residential area  $x_{12}$  and diagnosing hospital  $z_1$  are not significant, in line of the medical observations that (a) residential area is not as influential as gender  $x_{11}$  and age  $x_{13}$  and (b) the diagnosing hospital influences cancer outcomes only via cancer stages  $x_2$  diagnosed. Furthermore, patients with different cancer stages  $x_2 = 1, 2, 3, 4$  usually have rather different survivals, and so we model the mean  $\mu(x_{11}, x_{12}, x_{13}, z_1, x_2, z_2)$  separately for different cancer stages. Finally, we obtain

$$\begin{aligned}
 \mu(x_{11}, x_{13}, x_2 = 1, z_2) &= \beta_4 + \beta_5 x_{11} + \beta_6 x_{13} + \vartheta_{21,1} z_2 + \vartheta_{21,2} z_2 x_{13} \\
 \mu(x_2 = 2, z_2) &= \beta_7 + \vartheta_{22} z_2 \\
 \mu(x_2 = 3, z_2) &= \beta_8 + \vartheta_{23} z_2 \\
 \mu(x_2 = 4, z_2) &= \beta_9 + \vartheta_{24} z_2
 \end{aligned}
 \tag{2}$$

In the sub models for cancer stage  $x_2 = 2, 3, 4$ , gender  $x_{11}$  and age  $x_{13}$  are not significant in line with the medical observation that gender and age have less influence on cancer survival with advanced stages. For the cancer stage  $x_2 = 1$ , the point effect of  $z_2 = 1$  in stratum  $(x_{11}, x_{13})$  is  $\vartheta(x_{11}, x_{13}, x_2 = 1; 1) = \mu(x_{11}, x_{13}, x_2 = 1, 1) - \mu(x_{11}, x_{13}, x_2 = 1, 0) = \vartheta_{21,1} + \vartheta_{21,2} x_{13}$ . Its average over  $(x_{11}, x_{13})$  is the point effect of  $z_2 = 1$  for cancer stage  $x_2 = 1$  and is equal to  $\vartheta_{21,1} + \vartheta_{21,2} E(x_{13} | x_2 = 1)$ , denoted by  $\vartheta_{21}$ . For cancer stages  $x_2 = 2, 3, 4$ , the point effects of  $z_2 = 1$  are  $\vartheta(x_2; 1) = \mu(x_2, 1) - \mu(x_2, 0)$ , which are equal to  $\vartheta_{22}, \vartheta_{23}, \vartheta_{24}$ , respectively. The estimates  $\hat{\vartheta}_{21}, \hat{\vartheta}_{22}, \hat{\vartheta}_{23}$  and  $\hat{\vartheta}_{24}$  are presented in Table 2.

Briefly, we estimate the point effects of diagnosing and treating hospitals by modelling the means  $\mu(x_{11}, x_{12}, x_{13}, z_1)$  and  $\mu(x_{11}, x_{12}, x_{13}, z_1, x_2, z_2)$  in the usual framework of regression.

### 3.2 Blip effects of diagnosing and treating hospitals

Because neither (1) nor (2) contains the geographic area  $x_{12}$ , we do not include  $x_{12}$  in the following development. Here, we consider a special type of SCEs, called the blip effects, from which all other SCEs are determined,<sup>9,10</sup> see also the next subsection. The blip effect  $\phi(x_{11}, x_{13}; z_1)$  of diagnosing hospital  $z_1$  in the stratum  $(x_{11}, x_{13})$  is an

increase of the mean of potential survival when potentially and deterministically assigning a regime of diagnosing hospital  $z_1$  and small treating hospital versus a regime of small diagnosing hospital and small treating hospital to the stratum. Under the identifying condition, Robins' G-formula expresses the blip effect of  $z_1$  in terms of the means  $\mu(x_{11}, x_{13}, z_1, x_2, z_2)$  by<sup>1,2</sup>

$$\phi(x_{11}, x_{13}; z_1) = E\{\mu(x_{11}, x_{13}, z_1, x_2, 0)|x_{11}, x_{13}, z_1\} - E\{\mu(x_{11}, x_{13}, 0, x_2, 0)|x_{11}, x_{13}, 0\}$$

where the expectations are with respect to  $P(x_2|x_{11}, x_{13}, z_1)$  for  $z_1$  and  $z_1 = 0$ . Clearly, the blip effect of small diagnosing hospital  $\phi(x_{11}, x_{13}; 0) = 0$ . The blip effect  $\phi(x_{11}, x_{13}, z_1, x_2; z_2)$  of treating hospital  $z_2$  in the stratum  $(x_{11}, x_{13}, z_1, x_2)$  is an increase of the mean of potential survival when potentially and deterministically assigning treating hospital  $z_2$  versus 0 to the stratum. Under the identifying condition, Robins' G-formula gives<sup>1,2</sup>

$$\phi(x_{11}, x_{13}, z_1, x_2; z_2) = \mu(x_{11}, x_{13}, z_1, x_2, z_2) - \mu(x_{11}, x_{13}, z_1, x_2, 0)$$

Clearly, the blip effect of small treating hospital  $\phi(x_{11}, x_{13}, z_1, x_2; 0) = 0$ . Based on Robins' G-formula, one can estimate  $\phi(x_{11}, x_{13}; z_1)$  and  $\phi(x_{11}, x_{13}, z_1, x_2; z_2)$  by modelling  $\mu(x_{11}, x_{13}, z_1, x_2, z_2)$ .<sup>3,4</sup> However, in the model, both  $z_1$  and  $z_2$  are the variables of interest while  $x_2$  is a posttreatment variable of  $z_1$  and a confounder of  $z_2$ , so the modelling suffers from the curse of dimensionality and the null paradox.

An interesting clinical observation in cancer diagnosis is that young patients diagnosed at small hospitals tend to have poor prognoses.<sup>13</sup> This phenomenon is known as doctors' delay, but little-studied statistically. Motivated by this clinical observation, we suppose that  $\phi(x_{11}, x_{13}; 1)$  is a linear function of age  $x_{13}$ , that is, age is a modifier of the blip effect. We also studied gender  $x_{11}$  as an effect modifier, but found that the effect modification by gender was not significant at all. Further motivated by the point effect  $\vartheta(x_{11}, x_{13}, x_2; z_2)$  obtained in Section 3.1, where  $z_2$  is noticeably the last treatment variable, we suppose that the blip effects satisfy SNMM of the form

$$\begin{aligned} \phi(x_{11}, x_{13}; z_1) &= \gamma_{1,1}z_1 + \gamma_{1,2}x_{13}z_1 \\ \phi(x_{11}, x_{13}, z_1, x_2 = 1; z_2) &= \gamma_{21,1}z_2 + \gamma_{21,2}x_{13}z_2 \\ \phi(x_{11}, x_{13}, z_1, x_2 = 2; z_2) &= \gamma_{22}z_2 \\ \phi(x_{11}, x_{13}, z_1, x_2 = 3; z_2) &= \gamma_{23}z_2 \\ \phi(x_{11}, x_{13}, z_1, x_2 = 4; z_2) &= \gamma_{24}z_2 \end{aligned} \tag{3}$$

Let  $\gamma = (\gamma_{1,1}, \gamma_{1,2}, \gamma_{21,1}, \gamma_{21,2}, \gamma_{22}, \gamma_{23}, \gamma_{24})$  be the vector of all parameters in (3). The average of the blip effect  $\phi(x_{11}, x_{13}; 1)$  of  $z_1 = 1$  over  $(x_{11}, x_{13})$  is equal to  $\gamma_{1,1} + \gamma_{1,2}E(x_{13})$ , denoted by  $\gamma_1$ . Then  $\gamma_1 = \gamma_{1,1} + \gamma_{1,2}E(x_{13})$  is the blip effect of  $z_1 = 1$  in the population. The parameter  $\gamma_{1,2}$  is the modification of  $\gamma_1$  by  $x_{13}$  (age). The average of the blip effect  $\phi(x_{11}, x_{13}, z_1, x_2 = 1; 1)$  of  $z_2 = 1$  over  $(x_{11}, x_{13}, z_1)$  is equal to  $\gamma_{21,1} + \gamma_{21,2}E(x_{13}|x_2 = 1)$ , denoted by  $\gamma_{21}$ . Then  $\gamma_{21} = \gamma_{21,1} + \gamma_{21,2}E(x_{13}|x_2 = 1)$  is the blip effect of  $z_2 = 1$  for cancer stage  $x_2 = 1$ . The parameter  $\gamma_{21,2}$  is the modification of  $\gamma_{21}$  by  $x_{13}$ .

Now by applying the new G-formula<sup>10</sup> (see also formula (10) for  $t = 1$  with  $T = 2$  in Supplemental Data), we have

$$\begin{aligned} \vartheta(x_{11}, x_{13}; 1) &= \phi(x_{11}, x_{13}; 1) \\ &+ E\{\phi(x_{11}, x_{13}, z_1, x_2; z_2)|x_{11}, x_{13}, z_1 = 1\} - E\{\phi(x_{11}, x_{13}, 0, x_2; z_2)|x_{11}, x_{13}, z_1 = 0\} \end{aligned}$$

where the expectation is with respect to  $P(x_2, z_2|x_{11}, x_{13}, z_1)$  for  $z_1 = 1$  and  $z_1 = 0$ . This formula shows that  $\vartheta(x_{11}, x_{13}; 1)$  decomposes into the blip effects of  $z_1 = 1$  and  $z_2$ . Noticing that  $\phi(x_{11}, x_{13}, z_1, x_2; 0) = 0$ , we have

$$E\{\phi(x_{11}, x_{13}, z_1, x_2; z_2)|x_{11}, x_{13}, z_1\} = \sum_{j=1}^4 \phi(x_{11}, x_{13}, z_1, x_2 = j; 1)P(x_2 = j, z_2 = 1|x_{11}, x_{13}, z_1)$$

Let  $C_j(x_{11}, x_{13}) = P(x_2 = j, z_2 = 1|x_{11}, x_{13}, z_1 = 1) - P(x_2 = j, z_2 = 1|x_{11}, x_{13}, z_1 = 0)$ ,  $j = 1, 2, 3, 4$ , be the probability differences. Then under SNMM (3), we have

$$\vartheta(x_{11}, x_{13}; 1) = \gamma_{1,1} + \gamma_{1,2}x_{13} + (\gamma_{21,1} + \gamma_{21,2}x_{13})C_1(x_{11}, x_{13}) + \sum_{j=2}^4 \gamma_{2j}C_j(x_{11}, x_{13}) \tag{4}$$

The point effect  $\vartheta(x_{11}, x_{13}; 1)$  has been estimated from model (1) in Section 3.1. The probability differences  $C_j(x_{11}, x_{13})$  are estimated by the observed proportion differences  $\hat{C}_j(x_{11}, x_{13})$  in the data without modelling.

Because  $Z_2$  is the last treatment variable, we have  $\vartheta(x_{11}, x_{13}, x_2; 1) = \phi(x_{11}, x_{13}, z_1, x_2; 1)$ , where the point effect  $\vartheta(x_{11}, x_{13}, x_2; 1)$  has been estimated from model (2) in Section 3.1. Therefore, under SNMM (3), we have

$$\begin{aligned}\vartheta_{21,1} &= \gamma_{21,1} \\ \vartheta_{21,2} &= \gamma_{21,2} \\ \vartheta_{22} &= \gamma_{22} \\ \vartheta_{23} &= \gamma_{23} \\ \vartheta_{24} &= \gamma_{24}\end{aligned}\tag{5a}$$

From the estimates of the point effects, we can estimate the blip effects by applying (4) and (5a).<sup>10</sup> The point effects are not necessarily biased, and neither are those of the blip effects if SNMM (3) is not misspecified.<sup>10</sup>

Because age  $x_{13}$  is continuous and the number of its values may be large, it is difficult to directly use formula (4) to estimate the blip effects. On the other hand, the blip effects  $\phi(x_{11}, x_{13}; 1)$  are indexed by only two parameters  $\gamma_{1,1}$  and  $\gamma_{1,2}$  under SNMM (3). Hence, we may use the standardized point effects of  $z_1 = 1$  on a small number of subpopulations of patients at the time of diagnosis to estimate the blip effect.

As an illustration, let  $S_1 = L$  be the subpopulation with patients' age ( $x_{13}$ ) lower than the median age at the time of diagnosis and  $S_1 = H$  with age higher than or equal to the median age. Then, we have two standardized point effects of  $z_1 = 1$ , which are  $\Theta(S_1; 1) = E\{\vartheta(x_{11}, x_{13}; 1)|S_1\}$  with respect to the probability  $P(x_{11}, x_{13}|S_1)$  for  $S_1 = L, H$ . This procedure is known as the standardization in Epidemiology.

Let  $C_j(S_1) = E\{C_j(x_{11}, x_{13})|S_1\}$  ( $j = 1, 2, 3, 4$ ) be the mean of  $C_j(x_{11}, x_{13})$  with respect to  $P(x_{11}, x_{13}|S_1)$ . Then by taking the average of formula (4) with respect to  $P(x_{11}, x_{13}|S_1)$ , we obtain

$$\Theta(S_1; 1) = \gamma_{1,1} + \gamma_{1,2}E(x_{13}|S_1) + \gamma_{21,1}C_1(S_1) + \gamma_{21,2}E\{x_{13}C_1(x_{11}, x_{13})|S_1\} + \sum_{j=2}^4 \gamma_{2j}C_j(S_1), \quad S_1 = L, H \tag{5b}$$

The estimates  $\hat{E}(x_{13}|S_1)$ ,  $\hat{E}\{x_{13}C_1(x_{11}, x_{13})|S_1\}$ ,  $\hat{C}_j(S_1)$  ( $j = 1, 2, 3, 4$ ) are obtained, respectively, by averaging  $x_{13}$  with respect to  $\hat{P}(x_{13}|S_1)$ ,  $x_{13}\hat{C}_j(x_{11}, x_{13})$  with respect to  $\hat{P}(x_{11}, x_{13}|S_1)$ , and  $\hat{C}_j(x_{11}, x_{13})$  with respect to  $\hat{P}(x_{11}, x_{13}|S_1)$ , where all proportions involved are observed from data, not from modelling. Based on model (1) in the previous subsection, we have  $\vartheta(x_{11}, x_{13}; 1) = \vartheta_1$  and thus  $\Theta(S_1; 1) = E\{\vartheta(x_{11}, x_{13}; 1)|S_1\} = \vartheta_1$ . Therefore, we have  $\hat{\Theta}(S_1; 1) = \hat{\vartheta}_1$ , but the variance  $\text{var}\{\hat{\Theta}(S_1; 1)\}$  is obtained by adjusting  $\text{var}(\hat{\vartheta}_1)$  to the size of subpopulation  $S_1$ .

Now, conditional on all covariates, diagnosing and treating hospitals, we use (5a) and (5b) as a regression model to estimate  $\gamma = (\gamma_{1,1}, \gamma_{1,2}, \gamma_{21,1}, \gamma_{21,2}, \gamma_{22}, \gamma_{23}, \gamma_{24})$ , where the response variables are the estimates  $\hat{\Theta}(S_1; 1)$ ,  $\hat{\vartheta}_{21,1}$ ,  $\hat{\vartheta}_{21,2}$ ,  $\hat{\vartheta}_{2j}$  ( $j = 2, 3, 4$ ) while the explanatory variables are the proportions  $\hat{E}(x_{13}|S_1)$ ,  $\hat{E}\{x_{13}C_1(x_{11}, x_{13})|S_1\}$  and  $\hat{C}_j(S_1)$  ( $j = 1, 2, 3, 4$ ) and the one. These explanatory variables are obtained without modelling as describe above. The estimates  $\hat{\Theta}(S_1; 1) = \hat{\vartheta}_1$ ,  $\hat{\vartheta}_{21,1}$ ,  $\hat{\vartheta}_{21,2}$ ,  $\hat{\vartheta}_{2j}$  ( $j = 2, 3, 4$ ) are obtained in Section 3.1 from models (1) and (2), but their conditional variances are obtained from the same models by specifying a constant dispersion parameter for the distribution. The bootstrap method is used to obtain the covariance matrix  $\text{cov}(\hat{\gamma})$  incorporating the variability of all covariates and diagnosing and treating hospitals. With  $\hat{\gamma}$  and  $\text{cov}(\hat{\gamma})$ , we conduct the Wald test on  $\gamma$ .

Briefly, we estimate and test the blip effects of diagnosing and treating hospitals in the usual framework of regression. The result is presented in Table 2.

### 3.3 SCE under any regime of diagnosing and treating hospitals

First, we consider the SCE of treatment regime  $D_2$  in the stratum  $(x_{11}, x_{13}, z_1, x_2)$ :  $\text{SCE}(x_{11}, x_{13}, z_1, x_2; D_2)$ , which is an increase of the mean of potential survival when the regime  $D_2$  potentially and deterministically assigned treating hospital  $z_2$  versus 0 to the stratum. Noticeably, the regime  $D_2$  can be dynamic, for instance,  $D_2 = 1$  for patients younger than 69 years and  $D_2 = 0$  otherwise. According to the new G-formula<sup>10</sup> (see also formula (11) for time  $t = T = 2$  in Supplemental Data), we have  $\text{SCE}(x_{11}, x_{13}, z_1, x_2; D_2) = \phi(x_{11}, x_{13}, z_1, x_2; z_2)$ . Then we can readily use  $\phi(x_{11}, x_{13}, z_1, x_2; z_2)$  in SNMM (3) to obtain the following formula for the SCE under SNMM (3)

$$\text{SCE}(x_{11}, x_{13}, z_1, x_2; D_2) = \begin{cases} \gamma_{21,1}z_2 + \gamma_{21,2}x_{13}z_2, & x_2 = 1 \\ \gamma_{22}z_2, & x_2 = 2 \\ \gamma_{23}z_2, & x_2 = 3 \\ \gamma_{24}z_2, & x_2 = 4 \end{cases}\tag{6}$$

Second, we consider the SCE of treatment regime  $(D_1, D_2)$  in the stratum  $(x_{11}, x_{13})$ :  $SCE(x_{11}, x_{13}; D_1, D_2)$ , which is an increase in the mean of potential survival when the regime  $(D_1, D_2)$  potentially and deterministically assigned  $(z_1, z_2)$  of diagnosing and treating hospitals versus  $(0, 0)$  to the stratum. According to the new G-formula<sup>10</sup> (see also formula (11) for time  $t = 1$  with  $T = 2$  in Supplemental data), we have

$$SCE(x_{11}, x_{13}; D_1, D_2) = \phi(x_{11}, x_{13}; z_1) + E\{\phi(x_{11}, x_{13}, z_1, x_2; z_2)|x_{11}, x_{13}, z_1\}$$

where the expectation is with respect to  $P(x_2 = j|x_{11}, x_{13}, z_1)$ . This formula shows that the SCE decomposes into the blip effects of  $z_1$  and  $z_2$  in the regime. Under SNMM (3), we have

$$\begin{aligned} SCE(x_{11}, x_{13}; D_1, D_2) &= \gamma_{1,1}z_1 + \gamma_{1,2}x_{13}z_1 \\ &+ (\gamma_{2,1}z_2 + \gamma_{2,2}x_{13}z_2)P(x_2 = 1|x_{11}, x_{13}, z_1) + \sum_{j=2,3,4} \gamma_{2j}z_2P(x_2 = j|x_{11}, x_{13}, z_1) \end{aligned} \quad (7)$$

Formulas (6) and (7) imply that the SCEs are determined by blip effects.<sup>9,10</sup> Taking the average of (7) with respect to  $P(x_{11}, x_{13})$ , we obtain the formula for  $SCE(D_1, D_2)$ , which is an increase of the mean of the potential survival when  $(D_1, D_2)$  potentially and deterministically assigned  $(z_1, z_2)$  of diagnosing and treating hospitals versus  $(0, 0)$  to the population. In (7), the probability  $P(x_2|x_{11}, x_{13}, z_1)$  is modelled in the framework of single-point causal inference by assuming a parametric model and the multinomial distribution; see also a description in Supplemental Data for a general sequence of cancer diagnoses and treatments.

In (6) replacing  $\gamma = (\gamma_{1,1}, \gamma_{1,2}, \gamma_{2,1,1}, \gamma_{2,1,2}, \gamma_{2,2}, \gamma_{2,3}, \gamma_{2,4})$  by estimates  $\hat{\gamma}$  obtained from (5a) and (5b), we obtain the estimate for  $SCE(x_{11}, x_{13}, z_1, x_2; D_2)$  under any  $D_2$ . In (7) replacing probability  $P(x_2|x_{11}, x_{13}, z_1)$  by proportion  $\hat{P}(x_2|x_{11}, x_{13}, z_1)$  besides  $\gamma$ , we obtain the estimate for  $SCE(x_{11}, x_{13}; D_1, D_2)$  under any regime  $(D_1, D_2)$ . Averaging  $\widehat{SCE}(x_{11}, x_{13}; D_1, D_2)$  over the proportion  $\hat{P}(x_{11}, x_{13})$ , we obtain the estimate for  $SCE(D_1, D_2)$ . The variance of the estimated SCE is obtained by the bootstrap method. With the estimate and its variance, we obtain the confidence interval of the SCE. The covariance between the estimated SCEs under different regimes is also obtained by the bootstrap method. With the estimates and their covariance, we conduct the Wald test on the SCEs and so compare the underlying regimes.

By the dynamic programming procedure, we estimate the optimal regime  $(O_1, O_2)$  of diagnosing and treating hospitals by estimating and testing SCEs, as follows. Both  $D_1$  and  $D_2$  take values 0 or 1. From the estimates  $\widehat{SCE}(x_{11}, x_{13}, z_1, x_2; D_2)$ , we obtain  $\hat{O}_2 = \arg \max_{D_2} \widehat{SCE}(x_{11}, x_{13}, z_1, x_2; D_2)$ , which is the treating hospital such that  $\widehat{SCE}(x_{11}, x_{13}, z_1, x_2; \hat{O}_2)$  achieves the maximum value of  $\widehat{SCE}(x_{11}, x_{13}, z_1, x_2; D_2)$  for all possible  $D_2$  (0 or 1) given  $(x_{11}, x_{13}, z_1, x_2)$ . For the cancer stage  $x_2 = 1$ , we have according to (6)

$$\widehat{SCE}(x_{11}, x_{13}, z_1, x_2 = 1; D_2) = \hat{\gamma}_{2,1}z_2 + \hat{\gamma}_{2,2}x_{13}z_2$$

which, by using  $\hat{\gamma}_{2,1} = -0.801$  and  $\hat{\gamma}_{2,2} = 0.012$  in Table 2, achieves the maximum value when  $z_2 = 1$  for age  $x_{13} \geq 69$  years and when  $z_2 = 0$  for age  $x_{13} < 69$ . Thus, we have  $\hat{O}_2 = 1$  (large treating hospital) for  $x_{13} \geq 69$  years and  $\hat{O}_2 = 0$  (small treating hospital) otherwise. Here we see that  $\hat{O}_2$  is a dynamic regime. Let  $\Gamma_{21}$  be the average of  $SCE(x_{11}, x_{13}, z_1, x_2 = 1; \hat{O}_2)$  over  $x_{13}$ . For cancer stages  $x_2 = 2, 3$ , similarly, by applying (6) to the estimates  $\hat{\gamma}_{22} = 0.089$  and  $\hat{\gamma}_{23} = 0.027$  in Table 2, we find that  $\hat{O}_2 = 1$ . Let  $\Gamma_{22} = SCE(x_{11}, x_{13}, z_1, x_2 = 2; \hat{O}_2)$  and  $\Gamma_{23} = SCE(x_{11}, x_{13}, z_1, x_2 = 3; \hat{O}_2)$ . For the cancer stage  $x_2 = 4$  with  $\hat{\gamma}_{24} = -0.034$  in Table 2, we have  $\hat{O}_2 = 0$ . Let  $\Gamma_{24} = SCE(x_{11}, x_{13}, z_1, x_2 = 4; \hat{O}_2)$ , which is the baseline and equal to zero. Instead of  $\Gamma_{24} = 0$ , we present the estimate of  $\gamma_{24}$  in Table 2 together with the estimates of  $\Gamma_{21}$ ,  $\Gamma_{22}$  and  $\Gamma_{23}$ .

From the estimate  $\widehat{SCE}(x_{11}, x_{13}; D_1, \hat{O}_2)$ , we obtain  $\hat{O}_1 = \arg \max_{D_1} \widehat{SCE}(x_{11}, x_{13}; D_1, \hat{O}_2)$ , which is the diagnosing hospital such that  $\widehat{SCE}(x_{11}, x_{13}; \hat{O}_1, \hat{O}_2)$  is the maximum value of  $\widehat{SCE}(x_{11}, x_{13}; D_1, \hat{O}_2)$  for all possible  $D_1$  (0 or 1) given  $(x_{11}, x_{13})$ . Specifically, we replace  $D_2$  by  $\hat{O}_2$  and  $\gamma$  by its estimate  $\hat{\gamma}$  in (7) to get  $\widehat{SCE}(x_{11}, x_{13}; D_1, \hat{O}_2)$  and find that it achieves the maximum value when  $z_1 = 1$  for age  $x_{13} < 55$  years and when  $z_1 = 0$  for  $x_{13} \geq 55$ . Thus, we have  $\hat{O}_1 = 1$  (large diagnosing hospital) for  $x_{13} < 55$  and  $\hat{O}_1 = 0$  (small diagnosing hospital) for  $x_{13} \geq 55$ . Averaging  $\widehat{SCE}(x_{11}, x_{13}; \hat{O}_1, \hat{O}_2)$  over  $(x_{11}, x_{13})$ , we obtain the estimate for  $\Gamma_1 = SCE(\hat{O}_1, \hat{O}_2)$ , which is presented in Table 2.

Briefly, we use one estimated SNMM to estimate SCEs under various regimes and thus can compare these regimes by the hypothesis testing of the SCEs under these regimes.

### 3.4 Causal analysis of diagnosing and treating hospitals based on Table 2

First, we analyse treating hospital  $z_2$ . For the cancer stage  $x_2 = 1$ , the blip effect  $\gamma_{21}$  is estimated at 0.007 with  $P$ -value = 0.918, implying no overall difference between large and small treating hospitals. However, there is an effect modification

by age:  $\gamma_{21,2}$  estimated at 0.012 with  $P$ -value = 0.037. The estimated optimal treating hospital is the small one  $\hat{O}_2 = 0$  for age < 69 years and the large one  $\hat{O}_2 = 1$  for age  $\geq 69$  years. The corresponding optimal SCE,  $\Gamma_{21}$ , is estimated at 0.053 with  $P$ -value = 0.220. This observation reflects the fact that old patients usually have more comorbidities and large treating hospitals are probably better in dealing with comorbidities.

For the cancer stage  $x_2 = 2$ , the blip effect  $\gamma_{22}$  is estimated at 0.089 with  $P$ -value = 0.357, implying that a large treating hospital is better for patients with this cancer stage, albeit with somewhat small significance due to our small sample ( $n = 98$ ). There is no effect modification by age. The estimated optimal treating hospital is the large one ( $\hat{O}_2 = 1$ ) and the optimal SCE is equal to the blip effect:  $\Gamma_{22} = \gamma_{22}$ . Similar observations are made for  $x_2 = 3$ .

For the cancer stage  $x_2 = 4$ , the blip effect  $\gamma_{24}$  is estimated at  $-0.034$  with  $P$ -value = 0.382, implying that a small treating hospital is better for patients with an advanced cancer stage, albeit with somewhat small significance. There is no effect modification by age. The estimated optimal treating hospital is the small one ( $\hat{O}_2 = 0$ ) and the optimal SCE is equal to the baseline  $\Gamma_{24} = 0$ . This reflects the fact that palliative care is more central in patients with advanced stages of stomach cancer.

Second, we analyse the diagnosing hospital. The blip effect  $\gamma_1$  is estimated at  $-0.009$  with  $P$ -value = 0.710, implying no overall difference between large and small diagnosing hospitals. However, there is an effect modification by age:  $\gamma_{1,2}$  estimated at  $-0.001$  with  $P$ -value = 0.298. The estimated optimal diagnosing hospital is the large one ( $\hat{O}_1 = 1$ ) for age < 55 years and the small one ( $\hat{O}_1 = 0$ ) for age  $\geq 55$  years, implying that patients of age < 55 years benefited from large diagnosing hospitals. This reflects the delay in diagnosing stomach cancer among young patients at small hospitals, where cancer in young patients is very rare (a phenomenon called doctor's delay).

Third, we analyse the optimal regime in the population. Taking the average of  $SCE(x_{11}, x_{13}; \hat{O}_1, \hat{O}_2)$  over  $(x_{11}, x_{13})$ , we obtain the optimal SCE in the population,  $\Gamma_1$ , which is estimated at 0.028 with  $P$ -value = 0.237. This reveals the potential improvement for one-year survival by the optima regime.

Here, all effects are measured by the difference in mean survival, but they can also be measured by the difference in a function of mean survival. We may apply the new G-formula to estimate the causal effect measured as odds ratio, rate ratio and hazard ratio. Though the estimates are consistent, the problem of non-collapsibility of a non-linear measure becomes far worse in the context of sequential treatments, leading to biased estimates of these measures and false coverage probabilities of the confidence intervals for finite samples. Therefore, it is recommended that one uses the linear measure for the SCE.

In the medical example of this article, there was no censoring. Suppose there were non-informative censorings such as the loss of follow-up. Based on the new G-formula, we may address the loss of follow-up in the framework of single-point causal inference.<sup>9,10</sup> When estimating and testing the point effects and blip effects in Sections 3.1 and 3.2, we may simply remove the censored patients in the estimation. When estimating SCEs under various regimes in Section 3.3, we may simply remove the censored patients if the loss of follow-up occurred between  $Z_1$  and  $Z_2$ ; if the loss of follow-up occurs after  $Z_2$ , we may still use the data of these patients when estimating the covariate probability  $P(x_2|x_{11}, x_{13}, z_1)$  in (7).

## 4 Estimating and testing SCEs of cancer diagnosis and treatment on three-year survival

In Section 3, we replace one-year survival with three-year survival and follow the same procedure to estimate and test the blip effect and the SCE. The result is presented in Table 3.

Only for the cancer stage  $x_2 = 3$ , does the large treating hospital performs better than the small one with a considerable significance: the blip effect  $\gamma_{23}$  of treating hospital  $z_2 = 1$  is estimated at 0.069 with  $P$ -value = 0.168. For the cancer stage  $x_2 = 2$ , the blip effect  $\gamma_{22}$  of treating hospital  $z_2 = 1$  is estimated at  $-0.036$  with  $P$ -value = 0.728; due to the low significance and the medical knowledge, we believe that the large treating hospital performs better than the small one for  $x_2 = 2$ . Consequently, the estimated optimal treating hospital is the large one  $\hat{O}_2 = 1$  for all cancer stages  $x_2 = 1, 2, 3, 4$ . The estimated optimal diagnosing hospital is the large one  $\hat{O}_1 = 1$ . The SCE under the estimated optimal regime ( $\hat{O}_1, \hat{O}_2$ ),  $\Gamma_1$ , is estimated at 0.031 with  $P$ -value = 0.270.

## 5 Comparison of our method with available methods

Here, we only consider one-year survival. Method (i) is our method described in Section 3. Method (ii) is the parametric method based on Robins' G-formula.<sup>2-4</sup> Method (iii) is the marginal structural model based on the inverse probability of treatment weighting.<sup>4-6</sup> Method (iv) is the G-estimation based on SNMM or optimal-regime SNMM.<sup>2,4,7,8</sup>; this method incorporates both the Q-learning and A-learning in estimating optimal dynamic treatment regimes.

With our data, we aim to examine the modelling assumptions behind these methods and their abilities of estimating SCEs and comparing the underlying regimes of diagnosing and treating hospitals. These methods are active areas in the literature,<sup>4,8</sup> but to focus on the problems, we do not use advanced versions of these methods, which alleviate the



**Table 3.** Point effects, blip effects and optimal SCEs of diagnosing and treating hospitals on three-year survival of stomach cancer: estimate, *P*-value and 95% CI.

	Estimate, <i>P</i> -value and 95% CI in column for various causal effects				
	$\vartheta_1$	$\vartheta_{21}$	$\vartheta_{22}$	$\vartheta_{23}$	$\vartheta_{24}$
Point effect	0.024	0.011	-0.036	0.069	0.000
	0.413 (-0.034, 0.082)	0.882 (-0.132, 0.153)	0.728 (-0.242, 0.169)	0.168 (-0.029, 0.168)	0.997 (-0.029, 0.029)
Blip effect	$\gamma_1$ 0.014	$\gamma_{21}$ 0.011	$\gamma_{22}$ -0.036	$\gamma_{23}$ 0.069	$\gamma_{24}$ 0.000
	0.527 (-0.029, 0.058)	0.882 (-0.132, 0.153)	0.728 (-0.242, 0.169)	0.168 (-0.029, 0.168)	0.997 (-0.029, 0.029)
Optimal SCE	$\Gamma_1$ 0.031	$\Gamma_{21}$ 0.011	$\gamma_{22}$ -0.036	$\Gamma_{23}$ 0.069	$\gamma_{24}$ 0.000
	0.270 (-0.024, 0.086)	0.882 (-0.132, 0.153)	0.728 (-0.242, 0.169)	0.168 (-0.029, 0.168)	0.997 (-0.029, 0.029)

- Five point effects:  $\vartheta_1$  of large versus small diagnosing hospital;  $\vartheta_{2j}$  of large versus small treating hospital for cancer stage  $j = 1, 2, 3, 4$ .
- Five blip effects:  $\gamma_1$  of large versus small diagnosing hospital;  $\gamma_{2j}$  of large versus small treating hospital for cancer stage  $j = 1, 2, 3, 4$ .
- Five optimal SCEs:  $\Gamma_1$  under optimal regime of diagnosing and treating hospitals;  $\Gamma_{2j}$  of optimal treating hospital for cancer stage  $j = 1, 3$ .  $\Gamma_{24} = 0$  of optimal treating hospital for cancer stage 4 is the baseline, so the blip effect  $\gamma_{24}$  is instead presented. Similarly,  $\Gamma_{22} = 0$  and  $\gamma_{22}$  are presented.

**Table 4.** Comparison of our method with available methods in Section 5: estimate, *P*-value and 95% confidence interval (95% CI) for causal effects of diagnosing and treating hospitals on one-year cancer survival.

	Estimate, <i>P</i> -value and 95% CI in column of the causal effect			
	$\gamma_1$	$\gamma_{1,2}$	$Y_1$	$\Gamma_1$
Method (i)	-0.009	-0.001	-0.010	0.028
	0.710 (-0.054, 0.037)	0.298 (-0.003, 0.001)	0.580 (-0.059, 0.038)	0.237 (-0.019, 0.075)
Method (ii)	-0.340		-0.155	-0.012
	0.0 (-0.416, -0.265)		0.005 (-0.263, -0.048)	0.798 (-0.102, -0.079)
Method (iii)	0.290		-0.030	
	0.299 (-0.257, 0.836)		0.864 (-0.375, 0.315)	
Method (iv)	-0.009	-0.003	-0.011	
	0.713 (-0.054, 0.037)	0.322 (-0.009, 0.003)	0.662 (-0.059, 0.038)	

- Four estimation methods: method (i) our method; method (ii) the parametric method based on Robins' G-formula; (iii) the marginal structural model based on inverse probability of treatment weighting; method (iv) the G-estimation based on SNMM or optimal-regime SNMM; Empty cells imply that they are not easily estimable by the method.

- Causal effects:  $\gamma_1$  the blip effect of diagnosing hospital;  $\gamma_{1,2}$  the modification of  $\gamma_1$  by age;  $Y_1$  the optimal blip effect of diagnosing hospital;  $\Gamma_1$  the SCE under the optimal regime of diagnosing and treating hospitals described in Section 5.

problems. Methods (ii)–(iv) are also described in the context of this article in Supplemental Data. See also a simulation study for a comparison between these methods.<sup>10</sup>

Because methods (i)–(iv) may lead to the same inference of the causal effect of treating a hospital, we focus on the causal effect of diagnosing a hospital. In Section 3, we have applied method (i) to estimate the blip effect  $\gamma_1$ , its modification  $\gamma_{1,2}$  by age and the optimal SCE,  $\Gamma_1$ , on the population. Here, we additionally estimate the optimal blip effect of a large diagnosing hospital:  $Y_1 = SCE(D_1 = 1, \hat{O}_2) - SCE(0, \hat{O}_2)$ , because method (iv) does not estimate  $\Gamma_1$ . The result from these methods is presented in Table 4. From this table, we have the following observations.

With method (i), the modelling assumptions for the outcome are models (1) and (2) for estimating the point effects of  $z_1 = 1$  and  $z_2 = 1$  and SNMM (3) for estimating and testing the blip effects. When estimating and testing  $\gamma_1$ ,  $\gamma_{1,2}$ ,  $\Gamma_1$  and  $Y_1$ , a total of six parameters:  $\vartheta_1$  for the point effect of  $z_1 = 1$  and  $(\vartheta_{21,1}, \vartheta_{21,2}, \vartheta_{22}, \vartheta_{23}, \vartheta_{24})$  for the point effect of  $z_2 = 1$ , are involved.

Method (i) yields the estimates for  $\gamma_1 = -0.009$ ,  $\gamma_{1,2} = -0.001$ ,  $Y_1 = -0.010$  and  $\Gamma_1 = 0.028$  in line of the medical knowledge. It also yields the estimate for the optimal diagnosing hospital: a large one ( $\hat{O}_1 = 1$ ) for age  $< 55$  years and a small one ( $\hat{O}_1 = 0$ ) for age  $\geq 55$  years, which indicates the delay in diagnosing young patients at small hospitals (the phenomenon of doctor's delay). Method (i) allows for comparison between regimes ( $O_1, O_2$ ) and ( $0, 0$ ) by testing  $\Gamma_1$ . In general, it allows for comparison between different regimes, because their SCEs are estimated under the same modelling assumption.

With method (ii), we assume an unsaturated outcome model for the standard parameters  $\mu(x_{11}, x_{13}, z_1, x_2, z_2)$  as a trade-off between efficiency and bias, where  $x_{13}$  is categorized as lower than the median or higher. Even for this unsaturated model, a total of  $2 \times 2 \times 2 \times 4 \times 2 = 64$  standard parameters are involved in estimating and testing  $\gamma_1, \gamma_{1,2}, \Gamma_1$  and  $Y_1$ . Method (ii) yields the estimates  $\hat{\gamma}_1 = -0.340$ ,  $\hat{Y}_1 = -0.155$  and  $\hat{\Gamma}_1 = -0.012$ . The estimated optimal diagnosing hospital is the small one ( $\hat{O}_1 = 0$ ) for all ages. Method (ii) allows for comparison between regimes ( $O_1, O_2$ ) and ( $0, 0$ ) by testing  $\Gamma_1$ , because the means of the potential outcomes under the two regimes are estimated under the same model for  $\mu(x_{11}, x_{13}, z_1, x_2, z_2)$ . However, this method is not able to estimate  $\gamma_{1,2}$  due to the use of the unsaturated model. Furthermore, according to the medical knowledge, the estimates  $\hat{\gamma}_1$  and  $\hat{Y}_1$  cannot possibly be true as the differences in one-year survival between large and small diagnosing hospitals. The estimate  $\hat{\Gamma}_1$  is also biased, because the optimal SCE,  $\Gamma_1$ , should be larger than or equal to zero.

With method (iii), we assume models for the probabilities of assigning diagnosing and treating hospitals and use the estimated probabilities to calculate stabilized weight and non-stabilized weight for the outcome. With the stabilized weighted outcome, we estimate the mean of the potential outcome under a static regime and thus  $\gamma_1$ . With the non-stabilized weighted outcome, we estimate the mean of the potential outcome under a dynamic regime and thus  $Y_1$ . Method (iii) yields the estimate  $\hat{\gamma}_1 = 0.290$ . The estimate is not medically sensible possibly due to an imbalance between  $(z_1, z_2)$  and the covariates in the data.<sup>4-6</sup> Method (iii) also yields  $\hat{Y}_1 = -0.030$ . The estimated optimal diagnosing hospital is the small one ( $\hat{O}_1 = 0$ ) for all ages. Method (iii) is not able to estimate  $\gamma_{1,2}$  due to an imbalance between  $(z_1, z_2)$  and the covariates in the data.<sup>4,6</sup> Method (iii) does not estimate  $\Gamma_1$  either, which compares between the dynamic regime ( $O_1, O_2$ ) and static regime ( $0, 0$ ).

With method (iv), to estimate  $\gamma_1$  and  $\gamma_{1,2}$ , we assume SNMM (3) and a model for the mean of the potential outcome under regime ( $0, 0$ ), namely,  $E\{\mu(x_{11}, x_{13}, z_1 = 0, x_2, z_2 = 0)|x_{11}, x_{13}, z_1 = 0\}$  with respect to  $P(x_2|x_{11}, x_{13}, z_1 = 0)$  according to Robins' G-formula.<sup>1,2</sup> To estimate  $Y_1$ , we assume an optimal-regime SNMM and a model for the mean of the potential outcome under regime ( $0, \hat{O}_2$ ), namely,  $E\{\mu(x_{11}, x_{13}, z_1 = 0, x_2, z_2)|x_{11}, x_{13}, z_1 = 0\}$  with respect to  $P(x_2|x_{11}, x_{13}, z_1 = 0)$  according to Robins' G-formula, where  $\hat{O}_2$  deterministically and potentially assigned  $z_2$ .<sup>1,2</sup> Method (iv) yields nearly the same estimates for  $\gamma_1$  and  $\gamma_{1,2}$ , and  $Y_1$  as method (i), albeit with lower significance. The estimated optimal diagnosing hospital is the large one ( $\hat{O}_1 = 1$ ) for age  $< 68$  years and the small one ( $\hat{O}_1 = 0$ ) for age  $\geq 68$  years; it is different from the one obtained from method (i), possibly due to the misspecification of the model for the baseline. Method (iv) does not estimate  $\Gamma_1$ , so it is difficult to compare between regimes ( $O_1, O_2$ ) and ( $0, 0$ ). Generally, it is difficult to compare between different regimes, because their SCEs are estimated under different SNMMs. It is also difficult to specify models for the baselines.

## 6 Conclusion

In recent years, a huge amount of clinical data has become available, for instance, from various Swedish quality registers, which contain almost all economic and social information of a patient as well as a nearly complete record of clinical visits for the diagnoses and treatments of various cancers. Such data should lead to a large variety of comprehensive longitudinal studies of the influences of early diagnosis on various cancer outcomes such as survival and progression. In these studies, one needs not only to estimate SCEs under various regimes of diagnoses and treatments but also to compare these regimes.

In this article, we study the application of a parametric likelihood-based approach,<sup>10</sup> which allows for not only an unbiased estimation of SCEs under various regimes but also a comparison between these regimes by testing these SCEs under the same modelling assumption. Our method is implemented in three steps: first to estimate and test the point effect, second to use the estimated point effect to estimate and test the blip effect, and finally use the estimated blip effects to estimate and test SCEs. Each of these steps can be carried out using the usual regression and can be examined using the usual modelling tools, familiar to epidemiologists, in the causal inference of single-point treatment. Please note that the SCE is such a parameter that involves the entire data-generating mechanism of all covariates, treatments and the outcome across different diagnosing and treating stages, and consequently, it is highly difficult for a single procedure or algorithm to achieve both reliable estimation and hypothesis testing of the SCE.

Our medical example contains most of the essential components for evaluating the performance of cancer diagnosis: the blip effect, its modification by covariates and SCE under a general regime. It is also an observational study with continuous covariates. With its simple setting, we can readily extend this example to various cancer types; to different cancer outcomes such as cancer progression, quality of life and others; to different diagnosing techniques such as the biomarkers or a sequence of screening steps; to different modification factors such as social-economic status, comorbidity, and family

history. With complex covariate settings in cancer diagnoses, we believe that we may also apply more advanced methods than the usual regression, such as the targeted maximum likelihood method and even machine learning, to estimate the point effects.<sup>11</sup> If it is difficult to specify the distribution of the observable variables, we believe that we may also employ semi-parametric or non-parametric methods to estimate the point effects.

### Acknowledgements

Xiaoqin Wang and Li Yin were partially supported by Swedish Research Council with the grant number 2019 – 02913. All authors are grateful to the anonymous reviewer and editor for their comments and suggestions, which have considerably improved the article.

### Ethics approval

The proposed research is covered by the ethical committee approval (DNR880113/13, §121) from the ethical review board of Uppsala University.

### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.


### Author contribution

The four authors make equal contributions.

### Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Xiaoqin Wang and Li Yin were partially supported by the Swedish Research Council with the grant number 2019 – 02913.

### ORCID iD

Li Yin  <https://orcid.org/0000-0003-0410-8513>

### Supplemental material

(a) A general description of the new G-formula for the SCE in terms of point effects of treatments and a description of available methods (ii), (iii) and (iv) in the context of the medical example. (b) Data and code for all analyses of this article are available in [Zenodo], at <http://doi.org/10.5281/zenodo.4707971>.<sup>14</sup>

### References

1. Robins JM. A new approach to causal inference in mortality studies with sustained exposure periods – application to control of the healthy worker survival effect. *Mathematical Modeling* 1986; **7**: 1393–1512.
2. Robins JM. Causal inference from complex longitudinal data. In: Berkane M (eds) *Latent Variable modeling and applications to causality: lecture notes in statistics 120*. New York: Springer-Verlag, 1997, pp. 69–117.
3. Taubman SL, Robins JM, Mittleman MA, et al. Intervening on risk factors for coronary heart disease: an application of the parametric G-formula. *Int J Epidemiol* 2009; **2009**: 1599–1611.
4. Hernan MA and Robins JM. *Causal inference: what if*. Boca Raton: Chapman & Hall/CRC, 2020.
5. Hernan MA, Brumback B and Robins JM. Marginal structural models to estimate the causal effect of Zidovudine on the survival of HIV-positive men. *Epidemiology* 2000; **11**: 561–570.
6. Robins JM, Brumback B and Hernan MA. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000; **11**: 550–560.
7. Robins JM. Optimal structural nested models for optimal sequential decisions. In: Lin DY and Heagerty PJ (eds) *Proceedings of the second Seattle symposium in biostatistics, analysis of correlated data*. New York: Springer-Verlag, 2004, pp.189–326.
8. Kosorok MR and Laber EB. Precision medicine. *Annu Rev Stat Appl* 2019; **6**: 263–286.
9. Wang X and Yin L. Identifying and estimating net effects of treatments in sequential causal inference. *Electron J Stat* 2015; **9**: 1608–1643.
10. Wang X and Yin L. New G-formula for the sequential causal effect and blip effect of treatment in sequential causal inference. *Ann Stat* 2020; **48**: 138–160.
11. van der Laan MJ and Rubin DB. Targeted maximum likelihood learning. *Int J Biostatist* 2006; **2**: article 11.
12. Hansson LE, Ekström AM, Bergström R and Neren O. Surgery for stomach cancer in a defined Swedish population: current practices and operative results. *The European Journal of Surgery* 2000; **166**: 787–975.
13. Round T, Steed L, Shankleman J, Bourke L and Risi L. Primary care delays in diagnosing cancer: what is causing them and what can we do about them? *J R Soc Med* 2013; **106**: 437–440.
14. [All data and code] Yin L. 2021, Alternative G-formula to Estimate and Test Influence of Early Diagnosis on Cancer Survival.