Cause-specific or relative survival setting to estimate population-based net survival from cancer? An empirical evaluation using women diagnosed with breast cancer in Geneva between 1981 and 1991 and followed for 20 years after diagnosis

Robin Schaffar[1,2] *Bernard Rachet*[2], *Aurélien Belot*[2], *Laura Woods*[2]

1. Geneva Cancer Registry, Global Health Institute, University of Geneva, Geneva, Switzerland.

2. Cancer Research UK Cancer Survival Group, Department of Non-Communicable Disease Epidemiology, Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom

Corresponding author

Mr Robin Schaffar

Geneva Cancer Registry, Global Health Institute, University of Geneva,

55 Boulevard de la Cluse, 1205 Geneva, Switzerland.

+ 41.22.379.49.57

(robin.schaffar@unige.ch)

Abstract

Background.

Both cause-specific and relative survival settings can be used to estimate net survival, the survival that would be observed if the only possible underlying cause of death was the disease under study. Both resulting net survival estimators are biased by informative censoring and prone to biases related to the data settings within which each is derived. We took into account informative censoring to derive theoretically unbiased estimators and examine which of the two data settings was the most robust against incorrect assumptions in the data.

Patients and methods.

We identified 2,489 women in the Geneva Cancer Registry, diagnosed with breast cancer between 1981 and 1991, and estimated net survival up to 20-years using both cause-specific and relative survival settings, by tackling the informative censoring with weights. To understand the possible origins of differences between the survival estimates, we performed sensitivity analyses within each setting. We evaluated the impact of misclassification of cause of death and of using inappropriate life tables on survival estimates.

Results.

Net survival was highest using the cause-specific setting, by 1% at one year and by up to around 11% twenty years after diagnosis. Differences between both sets of net survival estimates were eliminated after recoding between 15% and 20% of the non-specific deaths as breast cancer deaths. By contrast, a dramatic increase in the general population mortality rates was needed to see the survival estimates based on relative survival setting become closer to those derived from cause-specific setting.

Conclusion.

Net survival estimates derived using the cause-specific setting are very sensitive to misclassification of cause of death. Net survival estimates derived using the relative-survival setting were robust to large changes in expected mortality. The relative survival setting is recommended for estimation of long-term net survival among patients with breast cancer.

Introduction

Net survival is defined as the survival that would be observed if the only possible underlying cause of death was the disease under study[1]. This definition of survival probability is of particular interest since it is not influenced by changes in mortality from other causes and therefore allows accurate evaluation of survival from the disease, essential for cancer control.

Two main approaches have been developed to estimate net survival, each requiring different data settings and assumptions. First, the cause-specific approach, which requires a data setting with reliable individual information on the underlying cause of death. Thus, only deaths from the cancer under study are defined as events whilst others are censored. Second, the relative survival approach[2] compares the overall survival of a cohort of patients to that which they would have experienced if they had had the same mortality experience of the general population from which they were drawn. This approach requires a different data setting, where mortality data about the population from which the cancer patients are drawn is available. Information about the cause of death is not required and we assume that the cancer-specific mortality included in the overall mortality is negligible compared to the overall mortality.

Both approaches are prone to a bias called *informative censoring*[3]. This is where the assumption of independence between the censoring process and the occurrence of the event (death) does not hold. For instance, an older patient is more likely to die from other causes than the disease under study than a younger patient. Thus, the older patients are more likely not to experience the death from cancer of interest simply because of their older age. The censoring process is therefore dependant on age and becomes informative. To take into account this bias, Robins[4] and Satten[5] proposed to weight the observed data by the inverse of the probability of not dropping out of the risk set, in order to find a cohort which would have been seen without the withdrawals. Pohar Perme[6] used this idea to propose an unbiased estimator of net survival within the relative survival setting.

As long as informative censoring is accounted for appropriately, both cause-specific and relative survival approaches derive theoretically unbiased estimators of net survival. However these estimators are prone to biases related to the data settings within which each is derived. These biases are independent of the method of estimation.

In the cause-specific data setting, what defines a cancer-related death versus a death from another cause is reliant upon the judgment of the person extracting the information and often prone to misclassification. Several studies have described this bias as being non-negligible[7–14]. For this reason the relative survival method has generally been preferred to estimate net survival with population-based data[15, 16]. However, within the relative survival data setting non-comparability between the cohort of patients and the general population[17] life tables used can also lead to bias. If a factor is differently distributed between patients and the general population, the resulting expected mortality of the cohort will be incorrectly estimated[18]. For instance, patients with lung cancer are more often smokers compared with the general population. Their expected mortality is therefore underestimated as they are more likely to die from other causes than the general population.[19] In the long term, this under-estimation may be balanced by the selection process over time of the more robust patients, who may die less than the general population[20]. This may impact net survival estimates[21]. Similarly, several factors can be associated with both cancer mortality and other diseases and lead to non-comparability between observed and expected mortality.

Our objective was to compare the two data settings, cause-specific and relative survival, when estimating long-term net survival. Both are subject to bias as described above; either misclassification of the cause of death or use of inappropriate life tables. We first derived theoretically unbiased estimators by using weights for both approaches, which took into account informative censoring. We then performed two sensitivity analyses in order to examine which of the two data settings was more robust against incorrect assumptions. We used each estimator as a reference for the other in order to evaluate the impact on the net survival estimates (Table 1).

We used data from the Geneva Cancer Registry which holds high quality data on cancer patients collected since 1970. This enabled us to evaluate the effect of these biases on long-term net survival. Furthermore, it afforded a privileged situation for estimating net survival within the cause-specific setting as information on cause of death had been independently verified.

Material and methods

Data

The data were provided by the Geneva Cancer Registry.

The Geneva Cancer Registry collects information on incident cancer cases from various sources, including hospitals, laboratories and private clinics, all requested to report new cancer cases. Trained registrars systematically extract information from the medical records and conduct further investigation in the case of missing key data. The registry regularly assesses survival, taking as the reference date the date of confirmation of diagnosis or the date of hospitalization (if it preceded the diagnosis and was related to the disease). In addition to passive follow-up (standard examination of death certificates and hospital records), active follow-up is performed yearly using the files of the Cantonal Population Office who maintain a register of the resident population. The cause of death is validated or revised from death certificates by registrars using all available clinical information. Autopsy reports, letter at death written by general practitioners and all patients' medical notes are used for the assessment of the revised cause of death. The treatment can therefore be considered as breast cancer death when information is found about it being part of the morbid events leading directly to death.[22] We included all women diagnosed with an invasive primary breast cancer between 1981 and 1991. These women have all been followed-up for a minimum of 20 years, and the last date of follow-up was 31st December 2011.

Statistical methods

Informative censoring

Informative censoring in a cohort of cancer patients is a differential selection process which affects the likelihood of the event of interest being observed. Different strategies have been derived for each data setting and are able to take into account informative censoring when estimating net survival (Appendix A).

The recently proposed Pohar-Perme[6] estimator enables informative censoring to be accounted for in the relative survival data setting, using weights calculated from the expected mortality of each cancer patient according to their individual characteristics. Expected mortality is derived from life tables for the general population from which the cancer patients are drawn and were previously smoothed.

In the cause specific setting, we used a similar strategy to weight the net survival estimator. We derived the weights using the cancer patient data and validated cause of death. We considered that the expected mortality of the cancer patients would be the same as the mortality rate from other causes of death than breast cancer amongst the cancer patients. We fitted a Poisson regression model to the cancer patient data where we considered death from a cause other than breast cancer as the event of interest. We adjusted on age at death and year of death. We used the model to derive expected mortality by age and year. We then used this set of rates to weight the breast-specific mortality hazard, in order to derive net survival estimates.

Potential biases related to the cancer data

In the relative survival setting, the potential bias of interest is related to the comparability between the cancer patients and the general population. An under-estimation of the expected survival would lead in an over-estimation of the net survival. On the contrary, an over-estimation of the expected survival would result in an under-estimation of net survival (Table 2).

In the cause-specific setting, the potential bias of interest is related to the accuracy of the classification of the cause of death. There are two possibilities; the proportion of breast cancer among the deceased patients is either over- or under-reported. If some non-specific deaths are misclassified as breast cancer deaths, the number of deaths from breast cancer is inflated. Net survival is therefore under-estimated. Similarly, net survival is over-estimated when some breast cancer deaths are misclassified as non-breast cancer deaths (Table 2).

Sensitivity analyses

In order to investigate the biases related to each data setting we performed two sensitivity analyses *(Table 3)*.

We defined the baseline situation as the estimation of net survival using the revised and/or validated cause of death in the cause-specific data setting, and the official Geneva life table in the relative survival data setting. We considered that both of these methods derive theoretically unbiased estimates of net survival.

We observed that in the baseline situation (Figure 1) net survival estimates derived using cause-specific data setting were higher than the relative survival data setting. We therefore concentrated our sensitivity analysis on two of the four potential biases to evaluate how this difference could have arisen (①and ④Table 2).

In scenario A, we evaluated whether the suitability of the life table in the relative survival data setting might be responsible for the difference observed (④Table 2). Mortality rates of the general population are only available by age, sex and calendar period in the Geneva canton. Nevertheless, other socio-demographic factors also influence the probability of death for an individual cancer patient. If the life table used in the relative survival setting does not accurately reflect the background risk of death of the cancer patient cohort, biased estimates of survival may result. This can happen, for example, because women with breast cancer tend to be more affluent than the overall population and these affluent women have a lower expected mortality rate than the population of women overall.

Deprivation information about cancer patients is available in three categories in the Geneva Cancer Registry (high, medium and low socio-economic position) but the expected mortality available is not detailed by deprivation. We therefore employed the rate ratios for the first, third and fifth quintiles of deprivation (derived from the England and Wales mortality rates for deprivation quintiles[23]) to build deprivation-specific expected mortality for the three socio-economic groups in the Geneva data. This generated a conservative situation because the differences between these quintiles in the English and Welsh data are likely to be greater than the (unknown) ratios between the three socio-economic groups in Geneva. This resulted in a substantially increased risk of death from other causes for the low socio-economic group and decreased risk for the high socio-economic group (Figure 2). These deprivation-specific life tables were then used to estimate net survival up to 20 years after diagnosis (scenario A1).

We further evaluated the degree to which expected mortality needed to increase in order to eliminate the difference we observed between the two estimators. We applied the life table for lowest socio-economic group (who have the highest mortality rates) to all the women in the data and estimated net survival up to 20 years after diagnosis (scenario A2).

We computed the difference between the baseline estimator of net survival using the cause-specific data setting and the relative survival estimators in scenario A1 and A2. Differences were smoothed by running a weighted non-parametric regression on time after diagnosis [24].

In scenario B, we considered misclassification of cause of death as the potential cause of the difference (① Table 2). Since the cause-specific data setting produced the higher of the two estimations, we considered only the situation in which breast cancer deaths had been misclassified as non-specific deaths. In this situation net survival calculated using the cause-specific approach would decrease.

We randomly re-attributed the cause of death variable from non-breast cancer to breast cancer for 10, 15, 20 and 25 per cent of the deceased patients (scenarios B1, B2, B3 and B4, respectively). We iterated this re-attribution 100 times and derived the mean cause-specific net survival up to 20 years for each scenario. The confidence interval was derived using the 95% coverage. The proportion of deaths due to breast cancer among deceased patients varied from 49.7% in the baseline situation to 62.2% in scenario B4 (table 4).

We computed the difference between the baseline estimator of net survival using the relative survival data setting and the cause-specific estimators for scenarios B1, B2, B3 and B4 respectively. Differences were smoothed by running a weighted non-parametric regression on time after diagnosis [24].

Results

The final cohort was comprised of 2,489 women diagnosed with an invasive breast cancer between 1981 and 1991 in Geneva, Switzerland.

Figure 1 shows the baseline situation where net survival estimator using the cause-specific setting was higher than the estimation using relative survival setting for all of the 20 years of follow-up.

The absolute difference between the two estimators increased with time after diagnosis from 1% at one year to 10.8% at 20 years. It remained less than 3% during the first ten years of follow-up (2.4% at 10 years) and started to increase more dramatically from 13 years onwards (Figure 1).

In scenario A1, where deprivation-specific life tables were applied, we observed a smaller but still substantial difference between the estimators (Figure 3). By contrast, the smoothed difference between the two different net survival estimators derived for scenario A2 (use of life tables of the most deprived population) was close to 0 during most of the first ten years after diagnosis (Figure 3).

Scenarios B1 to B4 correspond respectively to the re-allocation of 10, 15, 20 and 25% of deaths from non-breast cancer to breast cancer (Figure 4). As the proportion of re-allocation increased, the difference between the cause-specific approach and the baseline estimate derived within the relative survival data setting decreased, even turning negative. When 15-20% of the deaths were reallocated, the difference was close to zero. This suggested that with this level of reallocation the cause-specific approach and the relative survival approach used in the baseline situation derived a very similar estimate of net survival up to 10 years after diagnosis. Looking at results after 10 years, the two net survival estimators derived similar estimations when 25% of deaths were reallocated.

Discussion

Net survival is the survival that would be observed in a hypothetical world where the only possible underlying cause of death is the disease under study. This study is the first to account for informative censoring in the estimation of net survival in both cause-specific and relative survival data settings, allowing an accurate comparison of two unbiased estimators of net survival. Theoretically, both methods should give the same estimates of net survival. However, both net survival estimates are prone to biases related to the data and their specific assumptions. Differences in the estimates can be attributed to (i) incorrect expected mortality due to inadequate life tables in the relative survival data setting, or (ii) errors in the cause of death for some patients in the cause-specific data setting. We have evaluated these two possibilities using data on breast cancer patients whose cause of death has been independently validated, and who have been followed for 20 years after their diagnosis.

In the cause-specific setting, weights, estimated using the mortality hazard specific to the other causes of death from the cancer registry data, were applied to tackle informative censoring and estimate theoretically unbiased net survival. Although these internal weights were derived from a model, they may have been unstable due to small numbers of deaths in this fairly small breast cancer population. We thus also derived the weights using the expected mortality from the general population life tables (therefore similar as those used with the relative survival setting). Both weighting approaches gave very similar results, confirming the strength of the cause-specific setting estimator. We assumed that the weighted net survival estimator in the cause-specific setting was theoretically unbiased and equivalent to the Pohar Perme approach. However, simulation-based work is needed to assess its performance.

We observed that the cause-specific approach gave higher estimates of net survival compared to the relative survival approach (Figure 1). Moreover, the absolute difference between the two estimators increased with time since diagnosis up to 2.5% at 10 years after diagnosis and over 10% at 20 years. This enabled us to consider only two (① and ④, Table 2) out of the four potential biases related to the data setting. Indeed, we did not evaluate option ② or ③. Option ② considers the situation in which net survival is over-estimated because of under-estimated expected survival. This situation is unlikely insofar as cancer survivors are often prone comorbidities and are therefore no less likely to die than the general population, even if at longer term, the situation may be reversed, with the more robust patients being selected [25] [26]. Option ③ describes the situation in which net survival is under-estimated because deaths not due to breast cancer are mistakenly classified as breast cancer deaths. It is however more likely that the true number of breast cancer deaths is underestimated because deaths caused indirectly by breast cancer may be misclassified.

We first estimated net survival in the relative survival setting using different life tables to evaluate whether non-comparability between the general population and the cohort of patient under study compromises the estimation of net survival. In our study, we stratified the life table on deprivation (scenario A1) and the results showed that the net survival estimation was not substantially altered by this. In scenario A2, we noticed that a very large increase in the mortality rate was required (30% for a

50 year old woman) before the net survival estimated in the relative survival setting reached the net survival estimated in the cause-specific setting during the first 8 years after diagnosis. Such changes seem quite unreasonable at a population level. Net survival in the relative survival setting appeared therefore to be robust to inaccuracies in the underlying mortality rate.

The second sensitivity analysis showed that the level of misclassification could be relatively small to observe a large change in net survival. Indeed, recoding only 10% of the non-specific deaths led to a large decrease of the net survival using the cause-specific settings. The estimator of net survival derived with cause-specific method proved therefore to be relatively sensitive to the allocation of the cause of death. Recoding 15-20% of deaths from other causes to breast cancer resulted in the convergence of the net survival estimates in both settings. We have previously shown that a review of clinical files resulted in cause of death being revised for 9 per cent of women with breast cancer in the Geneva Cancer Registry data[22]. In the current study, 20% of the non-specific deaths represented 171 cases out of the 1700 deceased patients. As such, survival estimates are likely to be biased by the misclassification of a relatively small number of deaths that are indirectly related to breast cancer (for instance; side effects of treatment or suicide). The fact that the proportion of reallocation required to reduce the survival difference to zero increased after 10 years after diagnosis (from 10% to 15%) lends weight to this argument insofar that allocating breast cancer as an underlying cause of death is less probable with increasing time since diagnosis.

Taken together, these results suggest that survival derived with the cause-specific approach provides a very sensitive estimator, likely to be an overestimate of true net survival. On the contrary, survival derived with the relative survival approach is likely to be closer to the actual net survival of the patient cohort insofar as it is very robust to changes to the expected rate of death. This is especially true with increasing time after diagnosis.

Our study has considered only women with breast cancer. Breast cancer patients are not, however, representative of patients with cancer at different localisations. The proportions of specific deaths and age have a large impact in biases related to net survival. Future work will test the repeatability of our analyses on other cancer sites by different age groups. Preliminary results on cancers of colon-rectum, lung and on melanoma suggested results consistent with those provided by this study.

A dramatic increase in long-term survivors has been observed over the last few decades as a result of screening programs, more precise diagnostic tools and developments in treatment protocols.[27]. In the future a particular interest will be given to long-term net survival estimation, especially among younger patients.

Our results suggest that, when analysing routinely collected population-based data, the relative survival setting is likely to derive more accurate estimates of net survival, and that the cause specific setting is vulnerable to misclassification bias, particularly in the long-term. The relative survival setting is therefore highly recommended when estimating net survival with population-based data.

Table 1. Description of the two data settings available for the estimation of net survival.

Table 2: Potential biases related to data settings when estimating net survival.

Table 3. Description of the sensitivity analyses performed in order to check the extent of biases related to the data settings.

Figure 1. Net survival estimators in the baseline situation using both cause-specific and relative survival settings.

Figure 2. Geneva general population mortality rates by age for year 1991. Comparison between the baseline situation and scenarios A.

Table 4: Deaths distribution among female breast cancer patients diagnosed in Geneva between 1981 and 1991 according to scenarios.

Figure 3. Smoothed differences between the net survival estimators when informative censoring is taken into account: Cause-specific setting in the baseline situation vs. Relative survival setting in baseline situation, scenario A1 and scenario A2.

Figure 4. Smoothed differences between the net survival estimators when informative censoring is taken into account: Cause-specific setting in the baseline situation, scenarios B1, B2, B3 and B4 vs. Relative survival setting in the baseline situation.

Supplementary material

Appendix A. Description of net survival estimation when taking into account the theoretical bias of informative censoring.

Acknowledgements

This collaborative work was supported by the Geneva Cancer Registry and the Cancer Survival Group at London School of Hygiene and Tropical Medicine. We should like to thank Massimo Usel, Elisabetta Rapiti, Hyma Schubert, Christine Bouchardy and all collaborators at the Geneva Cancer Registry as well as Michel Coleman and all collaborators at the London School of Hygiene and Tropical Medicine for their support and encouragements.

Funding

This work was supported by the Swiss Cancer League [BIL KFS-3274-08-2013]

Disclosure

The authors have declared no conflict of interest.

References

1. Berkson J, Gage RP. Calculation of survival rates for cancer. 1952.

2. Ederer F, Cutler SJ. The relative survival rate: A statistical methodology. Natl. cancer Inst. Monogr. no. 6, 1961:101–121.

3. Danieli C, Remontet L, Bossard N et al. Estimating net survival: the importance of allowing for informative censoring. Stat. Med. 2012; 31(8):775–86.

4. Robins J. Information recovery and bias adjustemnt in proportional hazards regression analysis of randomized trials using surrogate markers. Proc. Biopharmmaceutical Sect. Am. Stat. Assoc. 1993.

5. Satten G. Estimating the marginal survival function in the presence of time dependent covariates. Stat. Probab. Lett. 2001; 54(4):397–403.

6. Perme MP, Stare J, Estève J. On estimation in relative survival. Biometrics 2012; 68(1):113–20.

7. Percy C, Stanek E, Gloeckler L. Accuracy of cancer death certificates and its effect on cancer mortality statistics. Am. J. Public Health 1981; 71(3):242–50.

8. Hoel DG, Ron E, Carter R. Influence of Death Certificate Errors on Cancer Mortality Trends Kiyohiko Mabuchi * Background : Evaluation of mortality data is an important tool in assessing Effects Research Foundation in. 1993; 85(13):1063–1068.

9. Messite J, Stellman SD. Accuracy of death certificate completion: the need for formalized physician training. JAMA 1996; 275(10):794–6.

10. Maudsley G, Williams EM. "Inaccuracy' in death certification--where are we now? J. Public Health Med. 1996; 18(1):59–66.

11. Lee PN. Comparison of autopsy, clinical and death certificate diagnosis with particular reference to lung cancer. A review of the published data. APMIS. Suppl. 1994; 45:1–42.

12. Lloyd-Jones DM, Martin DO, Larson MG, Levy D. Accuracy of death certificates for coding coronary heart disease as the cause of death. Ann. Intern. Med. 1998; 129(12):1020–6.

13. Newschaffer CJ, Otani K, Mcdonald MK, Penberthy LT. Causes of Death in Elderly Prostate Cancer Patients and. 21205:613–621.

14. Goldoni CA, Bonora K, Ciatto S et al. Misclassification of breast cancer as cause of death in a service screening area. Cancer Causes Control 2009; 20(5):533–8.

15. Crocetti E, De Lisi V, Gafà L et al. Net survival: comparison between relative and cause-specific survival estimates. Epidemiol. Prev. 2001; 25(3 Suppl):32–6.

16. Stelzner S, Hellmich G, Koch R, Witzigmann H. Exactitude of relative survival compared with cause-specific survival and competing risk estimations based on a clinical database of patients with colorectal carcinoma. Dis. Colon Rectum 2009; 52(7):1264–71.

17. Sarfati D, Blakely T, Pearce N. Measuring cancer survival in populations: relative survival vs cancer-specific survival. Int. J. Epidemiol. 2010; 39(2):598–610.

18. Baade PD, Fritschi L, Eakin EG. Non-cancer mortality among people diagnosed with cancer (Australia). Cancer Causes Control 2006; 17(3):287–97.

19. Ezzati M, Lopez AD. Estimates of global mortality attributable to smoking in 2000. Lancet 2003; 362(9387):847–52.

20. Zahl PH, Tretli S. Long-term survival of breast cancer in Norway by age and clinical stage. Stat. Med. 1997; 16(13):1435–49.

21. Ellis L, Coleman MP, Rachet B. The impact of life tables adjusted for smoking on the socioeconomic difference in net survival for laryngeal and lung cancer. Under Rev. 2014.

22. Schaffar R, Rapiti E, Rachet B, Woods L. Accuracy of cause of death data routinely recorded in a population-based cancer registry: impact on cause-specific survival and validation using the Geneva Cancer Registry. BMC Cancer 2014; In press(1):1–16.

23. Woods LM, Rachet B, Riga M et al. Geographical variation in life expectancy at birth in England and Wales is largely explained by deprivation. J. Epidemiol. Community Health 2005; 59(2):115–20.

24. Cleveland W. Robust locally weighted regression and smoothing scatterplots. J. Am. Stat. Assoc. 1979; 74:829–836.

25. Zahl PH. Frailty modelling for the excess hazard. Stat. Med. 1997; 16(14):1573–85.

26. Thomsen H, Brandt A, Sundquist J et al. Death causes in breast cancer patients. Ann. Oncol. 2012; (23):604–610.

27. Soerjomataram I, Louwman MWJ, Ribot JG et al. An overview of prognostic factors for long-term survivors of breast cancer. Breast Cancer Res. Treat. 2008; 107(3):309–30.

Table 1

			NET SURVIVAL	
		Setting	**Cause-specific**	**Relative survival**
Biases	**Theoretical / Methodological**		Informative censoring	
	Data		Misclassification of the cause of death	Non comparability between the cohort and the general population
Solutions	**Tackle informative censoring**	**Concept/Idea**	Weight the net survival estimator with the expected mortality	
		Application	Use the cancer data to estimate the expected mortality	Use the expected mortality derived from general population expected mortality
	Check the extent of biases related to the data	**Concept/Idea**	Sensitivity analyses: Modify the data to check the robustness of the net survival estimate	
		Application	Modify the number of specific death	Modify the expected mortality rates of the general population

Table 2

Net survival		Data setting: Cause-specific		Data setting: Relative-survival
Over-estimation	①	Real % of BCD 0  100 % of BCD in the data 0  100	②	Under-estimation of the expected survival $E_c < E_p$
Under-estimation	③	Real % of BCD 0  100 % of BCD in the data 0  100	④	Over-estimation of the expected survival $E_c > E_p$

BCD: Breast Cancer Deaths

E_p : Expected survival of the general population

E_c : Expected survival of the cancer patients

Table 3

	Setting	Baseline situation	A1	A2	B1	B2	B3	B4
			Scenario					
NET SURVIVAL	Cause-specific	Revised and/or validated cause of death	Revised and/or validated cause of death		Percentage of non-specific death reallocated: 10	15	20	25
	Relative survival	Official Geneva life table	Life table stratified by social class	Life table of the most deprived	Official Geneva life table			

Table 4

	Breast cancer death		Other cause of death		Total number of deaths	
	*N**	*%**	*N**	*%**	*N**	*%**
% of deaths reallocated†						
Real situation						
0%	844	49.7	856	50.4	1700	100.0
Scenario B1						
10%	930	54.7	770	45.3	1700	100.0
Scenario B2						
15%	972	57.2	728	42.8	1700	100.0
Scenario B3						
20%	1015	59.7	685	40.3	1700	100.0
Scenario B4						
25%	1058	62.2	642	37.8	1700	100.0

† Percentage of cases with non-specific cause of death randomly recoded as breast cancer.

**The numbers (N) and percentages (%) given are an average of the 100 iterations for scenarios B1, B2, B3 and B4.*

Appendix A

NET SURVIVAL

Net survival estimator	$\hat{\Lambda}_e(t) = \underbrace{\int_0^t \frac{dN(u)}{Y(u)}}_{\text{Estimated from the cancer data}} - \underbrace{\int_0^t \frac{\sum_{i=1}^{n} Y_i(u)\,d\Lambda_{P_i}(u)}{Y(u)}}_{\text{Estimated from the general population life table}}$		
Bias	Informative censoring		
Setting	**Cause-specific**		**Relative survival**
Event considered	Specific deaths		All deaths
Weights *W*	$X_i^W = \frac{X_i(t)}{S_{P_i}(t)}$, where $S_{P_i}(t)$ is the expected survival of individual *i* at time *t*		
Origin of the weights	Life table of the general population	Cancer data	Life table of the general population
Unbiased estimator of net survival	$\hat{\Lambda}_e(t) = \int_0^t \frac{dN_E^W(u)}{Y^W(u)}$		Pohar Perme estimator: $\hat{\Lambda}_e(t) = \int_0^t \frac{dN^W(u)}{Y^W(u)} - \int_0^t \frac{\sum_{i=1}^{n} Y_i^W(u)\,d\Lambda_{P_i}(u)}{Y^W(u)}$
Corresponding curve in figure 1	I	II	III

i	individual
t	time
W	weight
Y(t)	number of patients at risk up to t
N(t)	number of deaths up to t
$N_E(t)$	number of specific deaths up to t
$\Lambda_P(t)$	cumulative expected survival
$\Lambda_e(t)$	cumulative net survival

Figure 1
Click here to download high resolution image

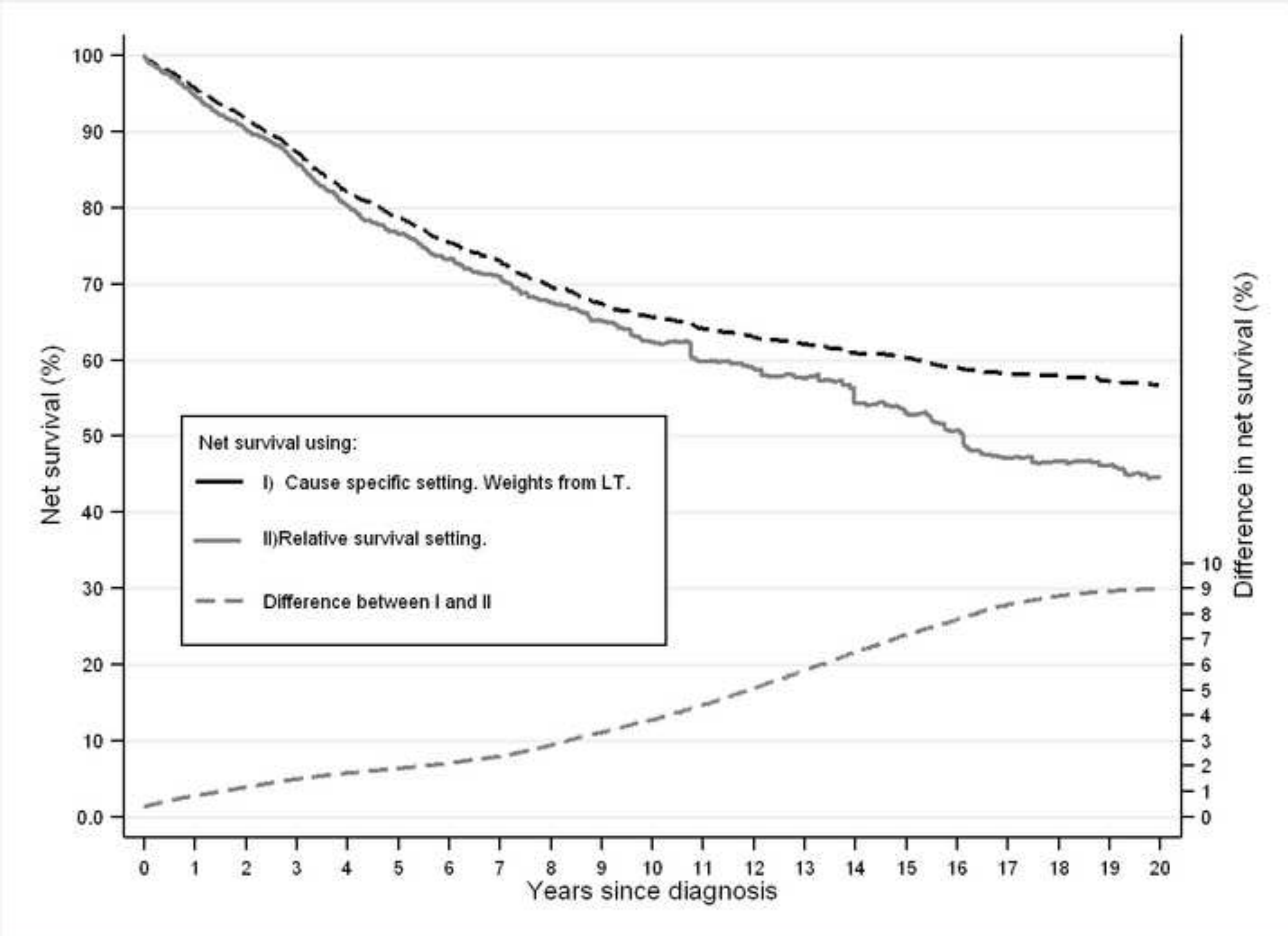

Figure 2
Click here to download high resolution image

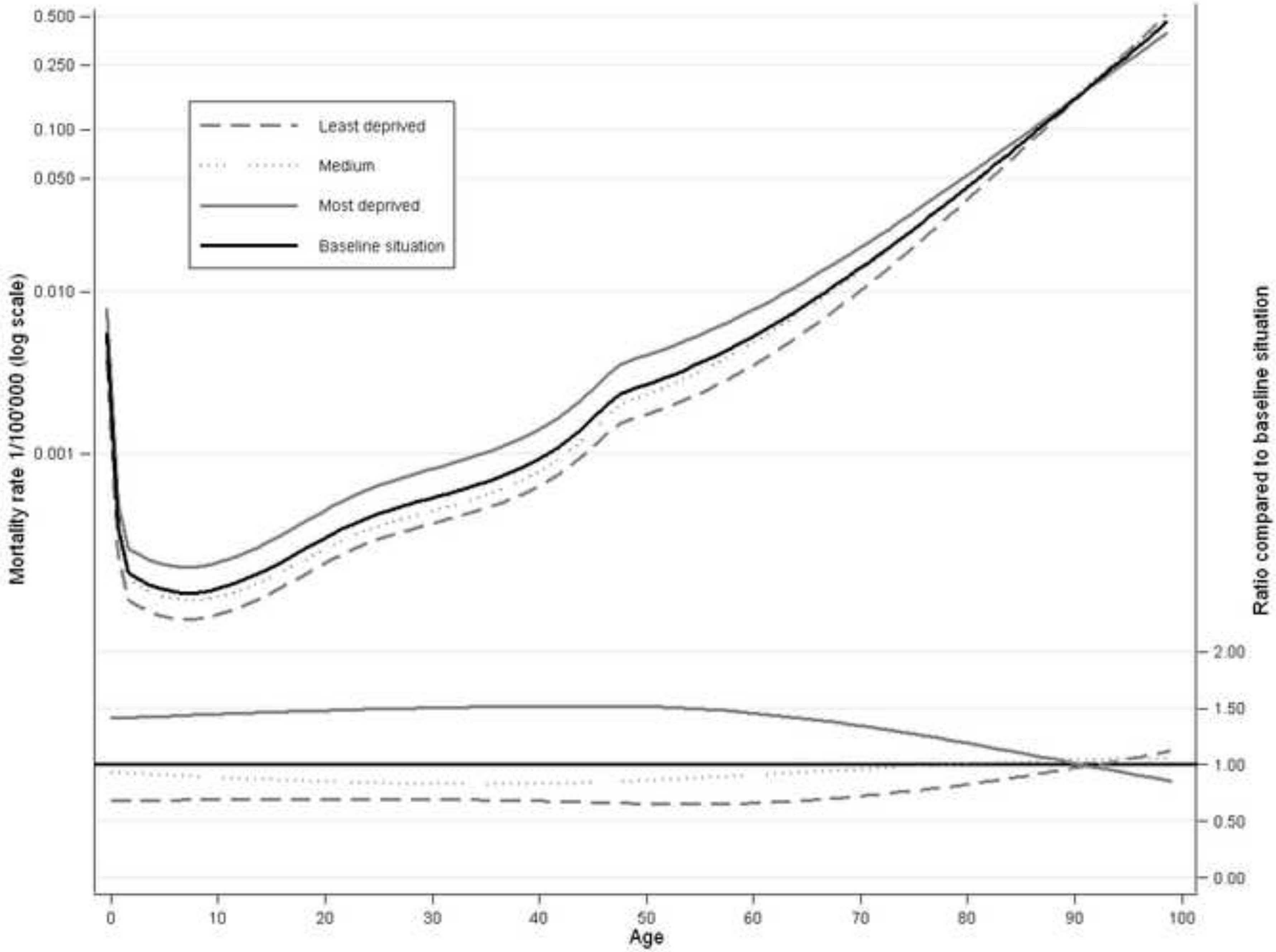

Figure 3
Click here to download high resolution image

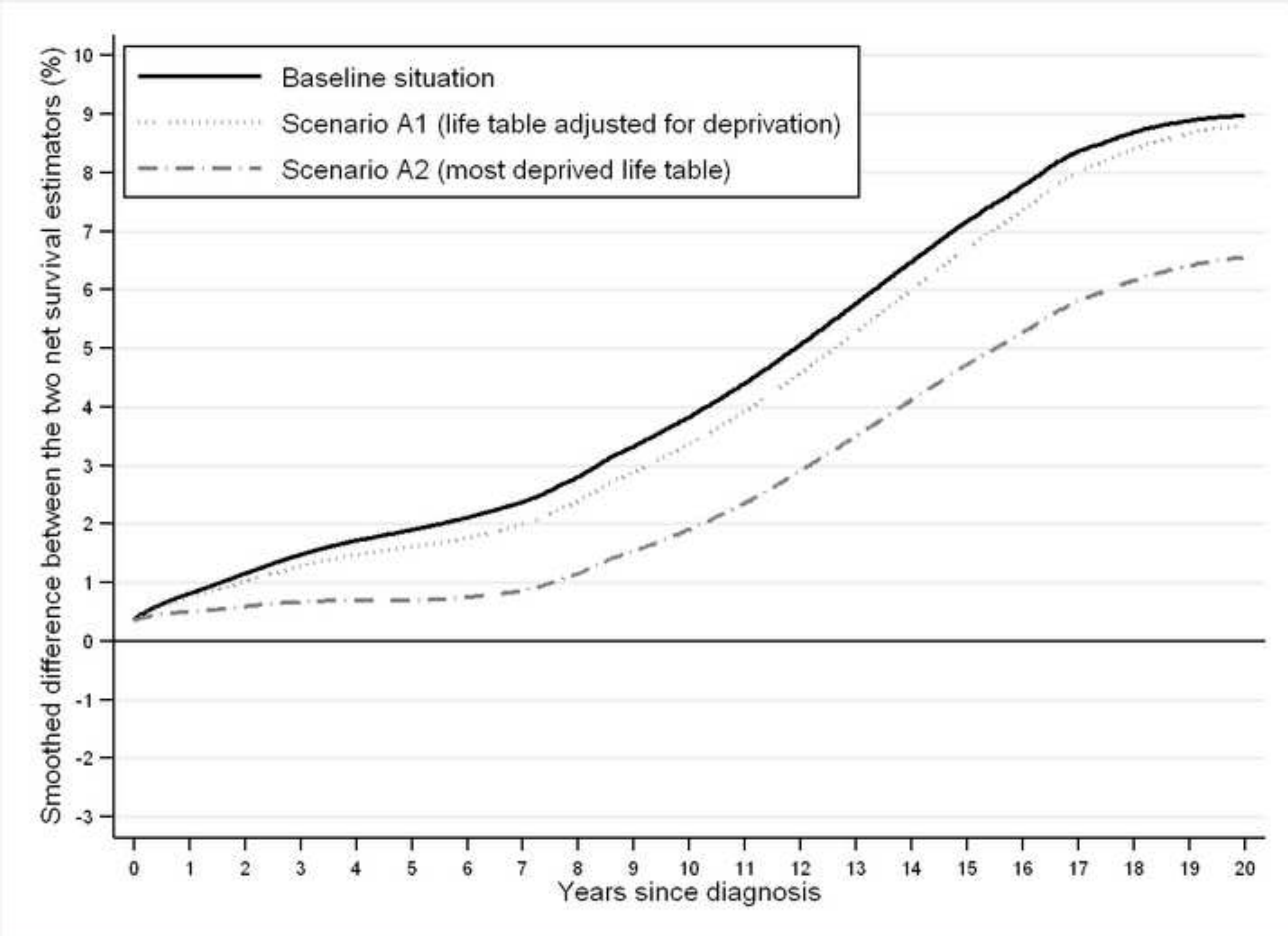

Figure 4
Click here to download high resolution image