Adjusting expected mortality rates using information from a control population: An example using socioeconomic status

Hannah Bower, Therese M-L Andersson, Michael J Crowther, Paul W Dickman, Mats Lambe and Paul C Lambert

Corresponding author: Hannah Bower, Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, 171 77 Stockholm, Sweden; email: hannah.bower@ki.se

Word count: 3996 (197 abstract)

Abstract

Expected or reference mortality rates are commonly used in the calculation of measures such as relative survival in population-based cancer survival studies and standardized mortality ratios. These expected rates are usually presented by age, sex and calendar year. In certain situations, stratification of expected rates by other factors is required to avoid potential bias if interest lies in quantifying measures by such factors, for example, socioeconomic status. If data on a population level are not available, information from a control population could be used to adjust expected rates. We present two approaches for adjusting expected mortality rates using information from a control population; a Poisson generalised linear model, and a flexible parametric survival model. We used a control group from BCBaSe, a Swedish register-based matched breast cancer cohort, to illustrate the two methods using socioeconomic status as a risk factor of interest. Our results showed that both the Poisson and flexible parametric survival approach estimate similar adjusted mortality rates by socioeconomic status. Additional uncertainty involved in the methods to estimate stratified expected mortality rates described in this study can be accounted for using a parametric bootstrap, but may make little difference if using a large control population.

Key words**:** Control group, expected mortality rate, flexible parametric model, life tables, Poisson model, socioeconomic factors, standard population, standardized mortality ratio.

Abbreviations: SMR, standardized mortality ratio; SES socioeconomic status; FPSM, flexible parametric survival model; SE, standard error.

Expected mortality rates calculated from a reference population are used in occupational, actuarial and medical research, and are often used as a comparative measure for an exposed population of interest. In order for estimates which use this comparative measure to be unbiased, it is important that expected mortality rates represent the mortality the exposed population would have experienced had they been unexposed (1-6). Thus, when one wishes to quantify the effects of factors which both the exposed population mortality and expected mortality rates vary by, expected mortality rates stratified by such factors should be used. Here we focus on population-based cancer studies where expected mortality rates are often presented in life tables by age, sex and calendar year. Relative survival (7,8), the loss in expectation of life (9) and standardized mortality ratios (SMRs) for cancers all use the expected mortality as a proxy for the mortality rate cancer patients would have experienced had they been cancer-free. It is of interest to consider how risk factors other than age, sex and calendar year affect cancer survival, for example, comorbidity or socioeconomic status (SES). Since both cancer-specific and all-cause mortality differ by SES (10,11), one must account for differences in mortality in the reference general population by SES if one wishes to estimate relative survival or similar measures by SES.

Previously this problem has been addressed by creating stratified life tables by using individualized information for the whole population (12-16). Unfortunately, it may not always be possible to obtain information for such factors for the whole population. There have also been some attempts to address this problem for SMRs in a Bayesian framework using different prior distributions (17). Others describe methods that create stratified life tables for such risk factors when individualized data are not available by using mortality information from subgroups of the reference population and prevalence of risk factors (4,12,18). We expand on these methods by presenting two models which use information from a control population to adjust expected mortality rates for covariates other than age,

sex and calendar period. The methods described here require a control population that accurately represents the reference population. This could either be a control population for the disease of interest, or another control population not related to the disease of interest. Using data on SES from a Swedish control population as an example, we also describe how to account for additional uncertainty involved in such methods by using parametric bootstrapping.

METHODS

Data

We used a cohort of breast cancer patients diagnosed between 1992 and 2012 in the Stockholm and Gotland regions of Sweden from a register-based database, BCBaSe, alongside matched controls; matching was based on age, year and region. Information on highest educational level was available for women with breast cancer and controls and used as a measure of SES. SES was categorised into low, medium and high; these corresponded to compulsory education less than or up to 9 years, up to 3 years of secondary education, and post-secondary education, respectively. Data on the Swedish population, obtained from Statistics Sweden (19) were used to create expected mortality rates specific to the Stockholm and Gotland regions by age and calendar year for females. Follow-up of control women began one-year post the diagnosis date of their matched breast cancer case to avoid a problem with the original matching that was based on the control being alive at the end of the year of diagnosis of the corresponding case woman. We restricted analyses to women with breast cancer and controls who were diagnosed between, or started follow-up between, 35 and 90 years of age. The study was approved by the Stockholm Regional Ethics Review Board.

Relative survival

We were interested in modelling relative survival of the breast cancer patients in BcBaSe by SES. Relative survival, $R(t)$, is defined as the ratio of the observed all-cause survival in the cancer patients, $S(t)$, and the survival expected had the cancer patients been cancer-free, $S^*(t)$; the latter of these is commonly taken as the expected survival from the general population, and t represents time. We were interested in relative survival as an estimator of net survival, the survival experienced by patients when they can only die of their cancer (20). Transforming the relative survival function to a

hazard function gives the excess hazard (or excess mortality rate), $\lambda(t)$. This is the difference between the all-cause mortality seen in the cancer patients, $h(t)$, and the mortality had the cancer patients been cancer free, $h^*(t)$. It is generally preferable to model the hazard function, but it is straightforward to transform back to the survival scale. We were interested in estimating the relative survival for the breast cancer patients by SES, age at diagnosis a_d and year of diagnosis y_d as a function of time since diagnosis, t:

$$R(t \mid \text{SES}, a_d, y_d) = \frac{S(t \mid \text{SES}, a_d, y_d)}{S^*(t \mid \text{SES}, a_d, y_d)}.$$

Standardized mortality ratio

In addition to relative survival, we were interested in calculating the SES-specific standardized mortality ratio (SMR). The SMR is the ratio of the observed number of deaths in the study population divided by the expected number of deaths in the reference population while standardizing over important covariates; the denominator is calculated by multiplying the mortality rate in the reference population with the person-time at risk in the study population (21). The reference population should represent the study population had they not been.

Methods to adjust expected mortality rates

We need $S^*(t \mid \text{SES}, a_d, y_d)$ in order to estimate the relative survival by SES as written above. On the hazard scale, this means we need the expected morality rates stratified by SES, age and year. Since we often only have expected mortality rates by sex, age and year from life tables, we were interested in estimating adjustment factors, ρ_j, for each SES group j; these can be functions of attained age

and attained year and are multiplied by the unadjusted mortality rate to give the SES-adjusted rate. The adjustment factors quantify the relative effect between the mortality rate in the control population, h_c, and the expected mortality rate in the whole population, h^* for each SES group. This can be formulated in the following way:

$$\frac{h_c(a,y,j)}{h^*(a,y)} = \exp\left[\sum_{j=1}^{3}\rho_j(a,y)\times \mathrm{SES}_j\right] \quad (1)$$

Two timescales, attained age a and attained year y are included. We use modelling approaches to utilize all information and avoid potential problems with small control populations. Note that sex is omitted from models in this application since the control population contained only females. In this paper we used a control population that was matched to our cases of interest, however, the described methodology could have been undertaken using a separate control population, provided that the control population was representative of the reference population. Similarly, the resulting SES-adjusted expected mortality rates calculated using the matched control population described here, could be used for analyses of diseases other than breast cancer patients.

Adjusting expected mortality rates using Poisson models. Equation 2 illustrates how to estimate adjustment factors ρ_j by using a generalised linear model framework with Poisson error and log link; d_c denotes the year-, age- and SES-specific death count and r_c the person-time at risk in the control population.

$$\ln(d_c(a,y,j)) = \sum_{j=1}^{3}\rho_j(a,y)\times \mathrm{SES}_j + \ln(h^*(a,y)) + \ln(r_c(a,y,j)) \quad (2)$$

Here $\ln(h_c(a,y,j)) = \ln\left(\frac{d_c(a,y,j)}{r_c(a,y,j)}\right)$. No baseline group was included in the Poisson model, instead

an offset of $\ln(h^*(a,y))$ was included so that covariate effects modelled deviations from this offset.

In order to model two timescales using this approach we split the timescale for every individual by attained age and attained year. We split in monthly intervals which increased the number of observations in our dataset from 133 361 to 29 553 897; we included the log of the original unadjusted expected rate, $\ln(\lambda^*(a,y))$, as an offset which assumed a constant rate for each year. Splitting time in this way can become more computationally intensive with increasing number of timescales and for smaller time increments.

Adjusting expected mortality rates using flexible parametric survival models. Flexible parametric survival models (FPSMs) can also be used to adjust expected mortality rates. FPSM (22) use restricted cubic splines (23) to model a transformation of the survival function. In this application, the models were used to model the log hazard function (24,25). A FPSM which models the log hazard function looks as follows:

$$ln(h(t;\mathbf{x})) = rcs(ln(t);\gamma_{\mathbf{0}}) + \mathbf{x}\beta + \sum_{d=1}^{D} rcs(ln(t);\gamma_{\mathbf{d}}) \times \mathbf{x_d}, \quad (3)$$

where t is time, $rcs(ln(t);\gamma_{\mathbf{0}})$ is a restricted cubic spline function of log time, D is the number of covariates with time-dependent effects, and $rcs(ln(t);\gamma_{\mathbf{d}})$ is the spline function for the d^{th} time-dependent effect.

In order to use a FPSM to estimate adjustment factors, we must smooth expected mortality rates and constrain coefficients in the FPSM to ensure there is no reference group represented by the baseline restricted cubic spline; this way covariate effects estimate deviations from the smoothed expected

rate. Expected mortality rates in life tables are empirical calculations of the rate in every age between 0 and 100 in yearly increments. These rates can fluctuate from year to year, particularly when considering a specific region in one country as in this application. This reasoning can be another argument for smoothing the expected mortality rates found in life tables; smoothing is also useful when abridged life tables are available (26). Smoothed expected mortality rates, h_s^*, were estimated by Poisson models using the number of deaths and person-time at risk from the original expected mortality rates, $d*$ and $r*$ respectively, in the following way:

$$\ln(h_s^*(a,y)) = \ln\left(\frac{d*(a,y)}{r*(a,y)}\right) = rcs_a(a,\gamma_a) + rcs_y(y,\gamma_y) \tag{4}$$

where attained age and attained year were modelled using restricted cubic spline functions rcs with four degrees of freedom. Here we assume that the relative effect of age is the same within each attained year, however, it is possible to extend this model by including interactions.

We use these smoothed expected mortality rates in FPSMs to estimate adjustment factors as follows:

$$\begin{aligned}\ln(h_c(a,y,j)) &= \underbrace{rcs_a(a;\gamma_a) + rcs_y(y;\gamma_y)}_{=\ln(h_s*(a,y))\text{in equation (4)}} + \sum_{j=1}^{3}\rho_j(a,y)\times \mathrm{SES}_j \\ &= rcs(ln(t);\gamma_0) + rcs_y(y;\gamma_y) + \sum_{j=1}^{3}\rho_j(a,y)\times \mathrm{SES}_j .\end{aligned} \tag{5}$$

The reference group represented by the spline function in the FPSM, $rcs(ln(t);\gamma_0)$, is constrained to have the same knot locations and same parameter values as the estimated age restricted cubic spline $rcs_a(a;\gamma_a)$ in equation 4. Two timescales are modelled simultaneously in equation 5 without the need for time-splitting as is required for the Poisson approach. The first timescale is modelled as the standard timescale in a FPSM by modelling the baseline log hazard function as a function of attained age. The second timescale, attained year, is incorporated by adding the year of birth of each individual to their attained age, calculating restricted cubic splines on this scale, and including in the model.

In both approaches, ρ_j was allowed to vary over attained age in a smooth manner but the effect of calendar year was assumed constant, i.e., the relative effect of age was assumed to be the same over all calendar years; two degrees of freedom were used for the splines included in the interaction between attained age and SES. The functional form assumed for the adjustment factors is highly important and should be checked using standard model checking methods, i.e., using AIC and/or BIC, and likelihood ratio tests. Sensitivity analyses of the functional form assumed here are presented in supplmetary materials.Stata code is also available as an online appendix which illustrates how to estimate adjustment factors using the two approaches.

Uncertainty in adjusted expected mortality rates. It is common to assume that there is no uncertainty in the expected mortality rates taken from life tables since these rates usually are based on the whole population. However, when creating adjusted expected mortality rates using a control population, there is uncertainty associated with the estimation of the adjustment factors that should be accounted for; this is a two-step process where the uncertainty of the estimated adjustment factors are not carried through to analyses which use the adjusted expected mortality rates. Accounting for this uncertainty can be done in the following way using a parametric bootstrap method:

1. Estimate adjustment factors using the control population
2. Draw from a multivariate-normal distribution using the mean vector **ρ** and the variance-covariance matrix from the model in step 1 to get new adjustment factors
3. Create new adjusted mortality rates using these new adjustment factors
4. Estimate relative survival and its variance using these new expected mortality rates
5. Save these estimates and repeat steps 2-4 a number of times, n

6. Estimate the adjusted-standard error, SE_{adjusted} by using Rubin's rules (27,28)

$$SE_{\text{adjusted}} = \sqrt{(\overline{W} + (1+1/n) \times B)}$$

where $\overline{W}$ is the mean of the variances estimated in each of the n bootstraps, and B is the variance of the relative survival estimates from the n bootsraps.

In this study, we use the above process with 100 repetitions where relative survival was estimated from flexible parametric survival models as described in the next section and the variance of the estimated relative survival was obtained using the delta method.

Relative survival and standardized mortality ratio analyses of Swedish breast cancer patients. 26 913 female breast cancer cases from BcBaSe were used to obtain estimates of SES-specific five-year relative survival. FPSMs were used to model the log excess cumulative hazard with three degrees of freedom. These models adjusted for SES, both age at diagnosis and year of diagnosis using restricted cubic splines with two degrees of freedom *using equally spaced quantiles of the distribution of the events*, included time-dependent effects for these three variables, and included an interaction between age at diagnosis and year of diagnosis. The expected mortality rates estimated from the unadjusted Swedish life table and SES-adjusted mortality rates by the methods described were used. We calculated confidence intervals for the relative survival estimates using a parametric bootstrap as described. We also calculated confidence intervals based on estimated adjustment factors using 10% and 1% random samples of the control population to assess how the confidence intervals vary using different sized control populations. Point estimates of relative survival taken from these different sample sizes were not directly comparable since only one random sample was selected; we were only interested in how the uncertainty in the relative survival varies.

We also estimated SMRs for patients diagnosed with stage I breast cancer by SES; stage I disease was chosen as an example to avoid presenting SMRs over all stages. We compared the SMRs when using the SES-adjusted and the original unadjusted expected mortality rates as the reference rate across several ages at diagnosis.

RESULTS

Descriptive statistics of the breast cancer cases and controls from the Stockholm and Gotland regions identified in the BCBaSe database and used in the analyses are shown in Table 1. For 3.7% of controls and 3.2% of cases SES data were missing (these data are not included in Table 1). 133 361 female controls were used for the adjustment of expected mortality rates and 26 913 female breast cancer cases for relative survival analyses after exclusions based on age at diagnosis. Figure 1 shows the estimated adjustment factors and Figure 2 the adjusted mortality rates from the two approaches. The adjustment factors were slightly different in the two methods, particularly for those younger than 50 years. Uncertainty was higher in the adjustment factors from both approaches where the mortality rates were smaller. Not surprisingly, the mortality rate for the high SES group was lower than the original unadjusted mortality rate. While the adjusted mortality rate for the medium SES group was similar to the unadjusted rate in those younger than 70 years, the adjusted mortality rate in the low SES was more similar to the unadjusted rate in those older than 70 years. Figure 2 also illustrates the effect of smoothing in the FPSM approach.

Figure 3 shows five-year relative survival by SES estimated from FPSMs using the original unadjusted life table mortality rates, and the SES-adjusted mortality rate from the FPSM approach. Similar results were seen for the Poisson model approach which are shown in Figure S2 in the supplementary materials. The relative survival was highest for the high SES group for all ages when using both the unadjusted and SES-adjusted expected mortality rate. Using the unadjusted expected mortality rates suggested that there was a larger difference in the five-year relative survival between the low and high SES groups. For example, the difference in the five-year relative survival between the low and high SES group when using unadjusted expected mortality rates in year 2000 for patients aged 45 years at diagnosis was 0.0521; this difference was 0.0266 when using SES-adjusted rates

from the FPSM. Table 2 displays SMRs using the unadjusted expected mortality rate, and the SES-adjusted expected mortality rate using the Poisson approach and the FPSM approach. There is a clear difference between the SMRs when using the unadjusted and SES-adjusted expected rates.

Confidence intervals for five-year relative survival were slightly widened due to the additional uncertainty of estimating the adjustment factors; see Figure 4 for high SES, and Figures S4 and S3 in the supplementary materials for medium and low SES, respectively. Tables S1-S6 in the supplementary materials show more detailed results for one-, five-, and ten-year relative survival. Figure 4 also shows confidence intervals for relative survival estimates when using 10% and 1% random samples of the control population. For those aged 45 at diagnosis in year 2000, the five-year relative survival was estimated at 0.9255 using adjusted expected mortality rates from the Poisson approach. The unadjusted and adjusted standard error for this estimate was 0.0617 and 0.0635, respectively. Larger differences in the adjusted confidence interval were seen in older ages, for example, the five-year relative survival estimated when using adjusted rated from the Poisson approach for 75 year olds in 2000 was 0.8843 with corresponding unadjusted and adjusted standard errors 0.0749 and 0.0801, respectively. When using only a random sample of 10% and 1% of the control population confidence intervals widened further; the five-year relative survival and the adjusted standard errors for the 10% and 1% samples were 0.0666 and 0.0955 respectively for patients aged 45 in year 2000.

DISCUSSION

Our results showed that both the Poisson and FPSM approaches can be used to adjust expected mortality rates and create adjusted life tables. Results from both methods yielded a higher mortality rate in the low SES group and a lower mortality rate in the high SES group. Results from relative survival analyses indicate that differences in mortality rates between the SES groups tend to be overestimated if SES-adjusted expected mortality rates are not used. Our findings corroborate results from earlier studies of socioeconomic gradients in breast cancer survival in Sweden (29-31) and previous findings of bias in estimates which used lifetables not adjusted for subgroups (4-6). We found little difference in the size of confidence intervals when using a parametric bootstrap in a large control population. Using only 1% of the control population, approximately 1 400 individuals, to estimate adjustment factors, resulted in larger confidence intervals. Because of a higher all-cause mortality in the older women there was a larger uncertainty in this group when estimating the adjustment factors, which translated into a wider confidence interval when accounting for this additional uncertainty. Any additional uncertainty in any relative survival estimates due to estimating adjustment factors may also reflect the complexity of the adjustment factors model.

We expected the adjusted mortality rate for the low SES group to be higher than the original unadjusted mortality rate. However, in women over 70 years the low-SES mortality rate was almost the same as the original unadjusted mortality rate; this similarity was due to a larger proportion of women this age being classified as low SES, particularly in earlier calendar years. This illustrates one potential problem of using educational level as a measure of SES, the opportunity to achieve a higher education was less common in earlier birth cohorts; suitable and stable measures of SES for use in research have been discussed extensively elsewhere (32,33). It may have also been useful to include additional covariates and interactions in the model to account for the difference in highest education

achieved over age and calendar year.

Both approaches for estimating adjustment factors are dependent on the control population being representative of the reference population i.e., transportability characteristics must hold (34). If the control population is not, then the adjusted mortality rates will be incorrect and analyses using these will be biased. For example, it is not advisable to use a male control group to estimate adjustment factors for a female population since the effect of SES on mortality rates may differ in males and females. However, it is possible for the adjustment factors estimated using the methods described here to be applied to different populations based on reasonable assumptions; for example, the adjustment factors estimated here for the Stockholm and Gotland regions are likely to be applicable to the whole of Sweden. In this illustration, a matched control population was used to estimate adjustment factors and adjusted expected mortality rates. It is important to note that the control population used in the estimation of the adjustment factors need not be a control population which was previously matched to the diseased population of interest..

We saw minor differences in the adjustment factors and their confidence intervals between the Poisson and the FPSM. The reason for these differences could be due to 1) differences when considering continuous time, as in the FPSM approach, versus assuming a constant expected mortality within yearly intervals, as in the Poisson approach; or 2) the effect of using smoothed rates plus the model chosen to smooth the expected mortality rates. It is possible to use the Poisson approach when using smoothed expected mortality rates, thus we repeated analyses using this method and saw closer agreement between the adjustment factors estimated from this model and those estimated from the FPSM. However, small differences were seen in the absolute adjusted rates between the two methods since the largest differences in adjustment factors were seen in those younger than 50 years where the

absolute rate is very small. In the same manner, one possible limitation of the smoothing implemented in this study is that the model used was simple. While it is possible to use a more complex smoothed rates model, sensitivity analyses illustrated that the smoothed rates model offered a reasonable fit; Figure S1 in the supplementary materials shows the fit of different smoothed rates models for year 2000. On the other hand, it may not be reasonable to assume a constant expected mortality rate within a yearly period, and smoothing offers an alternative to this assumption.

This work expands on previous work using SMRs to create sub-population, or adjusted, expected mortality rates. These previous methods require information on the prevalence of exposures in the whole population (4,18), or use covariate-specific SMRs (12), e.g., for a specific age, year and SES group. Although these methods also estimate adjusted expected mortality rates, our methods have additional advantages. Firstly, information on the whole reference population is not required; we instead suggest using information often recorded in control populations, related or unrelated to the diseased population, and make the assumption that the control population is representative of the reference population. Here, we assumed the stratified mortality rates (by SES, age and year) in the control population were representative of the reference population. Secondly, it is advantageous to use both information from the control population and information from the general (reference) population. Thirdly, covariate-specific SMRs are usually available for categorised covariates, here we are able to capture a continuous effect of covariates using a modelling framework.

Conclusions

Both the Poisson and FPSM approaches can be used to adjust expected mortality rates using a control population; these adjusted rates are essential for avoiding biased results in certain situations when calculating relative survival, SMRs, and other similar measures. Since the Poisson model is a

generalized linear model and available in different statistical software, it will be the more familiar approach to many. The ability to model two timescales simultaneously in the FPSM approach is highly useful to avoid long computational time when using a large dataset. Accounting for the additional uncertainty involved in estimating adjustment factors using a control population can be done using parametric boostrapping, and should be done for small control populations, but may not make a large difference when the control population is large. Since data on factors such as SES are often more easily available in control populations, the methodology presented in this study could offer the possibility of estimating relative survival, SMRs, and other measures to answer clinically relevant questions for risk factors other than sex, age and calendar year of diagnosis.

Acknowledgements

Author affiliations: Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden (Hannah Bower, Therese M-L Andersson, Paul W Dickman, Mats Lambe and Paul C Lambert); Department of Health Sciences, University of Leicester, Leicester, UK (Michael J Crowther, Paul C Lambert); and Regional Cancer Center, Uppsala University Hospital, Uppsala, Sweden (Mats Lambe).

This work was supported by grants from the Swedish Research Council and the Swedish Cancer Society.

Conflict of interest: none declared.

References

1. Rothman KJ, Greenland S, eds. *Modern Epidemiology*. 2nd ed. Philadelphia, PA: Lippincott-Raven, Publishers, 1998.

2. Morfeld P, Büchte SF, McCunney RJ, et al. Lung cancer mortality and carbon black exposure: Uncertainties of SMR analyses in a cohort study at a German carbon black production plant. *Journal of Occupational Health and Environmental Medicine*. 2006; 48:1253–1264.

3. Chaturvedi AK, Mbulaiteye SM, Engels EA.Underestimation of relative risks by standardized incidence ratios for AIDS-related cancers. *Annals of Epidemiology*. 2008; 18:230–234.

4. Sarfati D, Hill S, Blakely T, et al. The effect of comorbidity on the use of adjuvant chemotherapy and survival from colon cancer: a retrospective cohort study. *BMC Cancer*.2009; 9:116.

5. Blakely T, Soeberg M, Carter K, et al. Bias in relative survival methods when using incorrect life-tables: lung and bladder cancer by smoking status and ethnicity in New Zealand. *Int J Cancer*.2012; 131):E974–E982.

6. Graffeo N, Jooste V, Giorgi R. The impact of additional life-table variables on excess mortality. *Statistics in Medicine*. 2012; 31:4219–30.

7. Ederer F, Axtell LM, Cutler SJ. The relative survival rate: A statistical methodology. *National Cancer Institute Monograph*. 1961; 6:101–121.

8. Dickman PW, Adami H-O. Interpreting trends in cancer patient survival. *J Intern Med*. 2006; 260(2):103–117.

9. Andersson TM-L, Dickman PW, Eloranta S, et al. Estimating the loss in expectation of life due to cancer using flexible parametric survival models. *Statistics in Medicine*. 2013; 32:5286–5300.

10. Kogevinas M, Marmot MG, Fox AJ, et al. Socioeconomic differences in cancer survival. *Journal of Epidemiology and Community Health*. 1991; 45:200–203.

11. Feinstein JS. The relationship between socioeconomic status and health: A review of the literature. *Millbank Quarterly*.1993; 71:279–322.

12. Dickman PW, Auvinen A, Voutilainen ET, et al. Measuring social class differences in cancer patient survival: is it necessary to control for social class differences in general population mortality? A Finnish population-based study. *J Epidemiol Community Health*. 1998; 52(11):727–734.

13. Jeffreys M, Stevanovic V, Tobias M, et al. Ethnic inequalities in cancer survival in New Zealand: linkage study. *American Journal of Public Health*. 2005; 95:834–7.

14. Eloranta S, Lambert PC, Cavalli-Björkman N et al. Does socioeconomic status influence the prospect of cure from colon cancer–a population-based study in Sweden 1965-2000. *Eur J Cancer*. 2010; 46(16):2965–2972.

15. Morris M, Woods LM, Rachet B. A novel ecological methodology for constructing ethnic-majority life tables in the absence of individual ethnicity information. *Journal of epidemiology and community health*. 2015; 69:361–7.

16. Rachet B, Maringe C, Woods LM, et al. Multivariable flexible modelling for estimating complete, smoothed life tables for sub-national populations. *BMC Public Health*. 2015; 15:1240.

17. Morfeld P and McCunney RJ. Bayesian bias adjustments of the lung cancer SMR in a cohort of German carbon black production workers. *Journal of Occupational Medicine and Toxicology*. 2010;5:23.

18. Ellis L, Coleman MP, Rachet B. The impact of life tables adjusted for smoking on the socio-economic difference in net survival for laryngeal and lung cancer. *Br J Cancer*. 2014; 111(1):195–202.

19. Statistiska centralbyrån (Statistics Sweden). Statistical database. http://www.statistikdatabasen.scb.se/pxweb/en/ssd/?rxid=c5ce91b5-9830-4994-93a7-dd9539464d9a

Accessed Dec 2016.

20. Dickman PW, Coviello E. Estimating and modelling relative survival. *The Stata Journal*,. 2015; 15(1):186–215.

21. Richardson DB, Keil AP, Cole SR et al. Observed and Expected Mortality in Cohort Studies. *American Journal of Epidemiology*.2017; published online ahead of print

22. Royston P, Parmar MKB. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine*.2002; 21(15):2175–2197.

23. Durrleman S, Simon R. Flexible regression models with cubic splines. *Statistics in Medicine*. 1989;8(5):551–561.

24. Crowther MJ, Lambert PC. A general framework for parametric survival analysis. Statistics in Medicine. 2014;33(30):5280-5297

25. Bower H, Crowther MJ, Lambert PC. strcs: A command for fitting flexible parametric survival models on the log-hazard scale. The Stata Journal. 2016;16(4):989-1012.

26. Baili P, Micheli A, Montanari A, et al. Comparison of four methods for estimating complete life tables from abridged life tables by using mortality data supplied to EUROCARE-3. *Math Popul Stud*. 2005; 12:183–98.

27. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. Hoboken, NJ: John Wiley and Sons, Publishers, 2004.

28. Marshall A, Altman DG, Holder RL, et al. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. *BMC Medical Research Methodology*. 2009: 9:1–8.

29. Lagerlund M, Bellocco R, Karlsson P, et al. Socio-economic factors and breast cancer survival - a population-based cohort study (Sweden). *Cancer Causes and Control*. 2005;16:419–430.

30. Halmin M, Bellocco R, Lagerlund M , et al. Long-term inequalities in breast cancer survival - a ten year follow-up study of patients managed within a National Health Care System (Sweden). *Acta Oncologica*. 2008;47:216–224.

31. Hussain SK, Altieri A, Sundquist J, et al. Influence of education level on breast cancer risk and survival in Sweden between 1990 and 2004. *International Journal of Cancer*. 2008; 122:165–169.

32. Braveman PA, Cubbin C, Egerter S, et al. Socioeconomic status in health research: one size does not fit all. *Jama*. 2005; 22:2879–88.

33. Shavers VL. Measurement of socioeconomic status in health disparities research. *Journal of the National Medical Association*,. 2007;99:1013–23.

34. Hernán MA, Robins JM. *Causal Inference*. Boca Raton: Chapman & Hall/CRC, forthcoming. Accessed May 23, 2017.

Figures

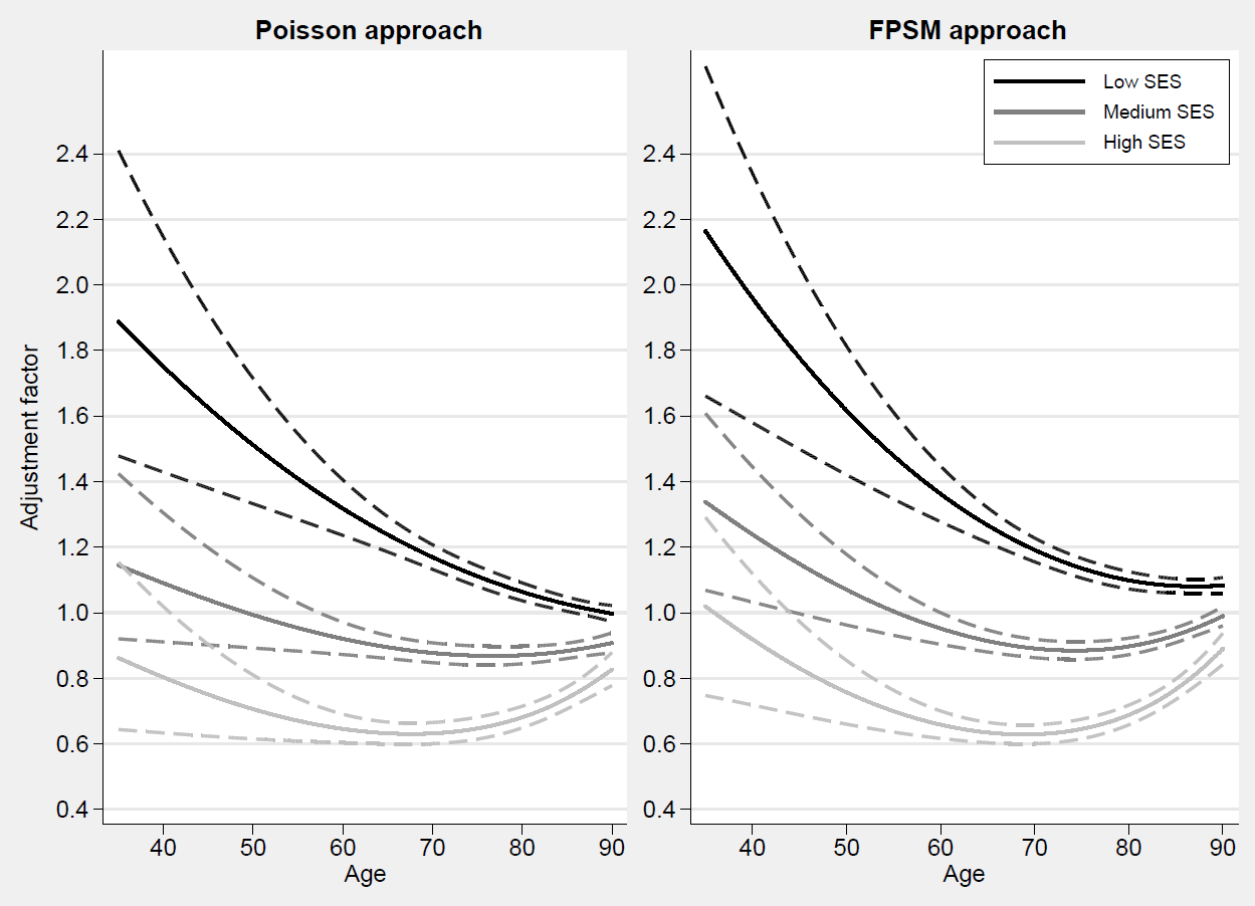

Figure 1: Adjustment factors (solid lines) with 95% confidence intervals (dashed lines) for each socioeconomic (SES) group from the Poisson and flexible parametric survival model (FPSM) approaches; Stockholm and Gotland regions of Sweden in year 2000

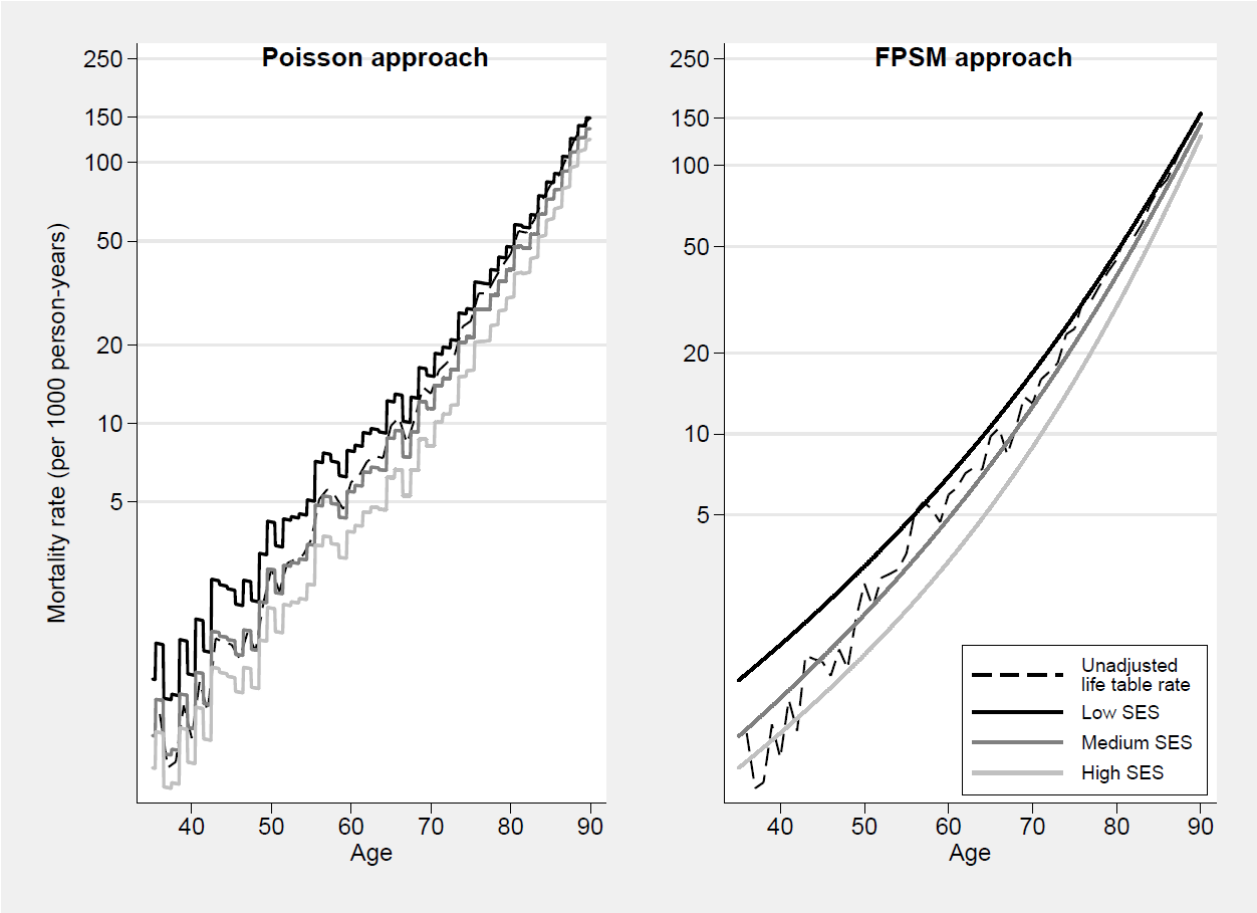

Figure 2: Adjusted mortality rates for each socioeconomic (SES) group from the A) Poisson and B) flexible parametric survival model (FPSM) approaches; Stockholm and Gotland regions of Sweden in year 2000

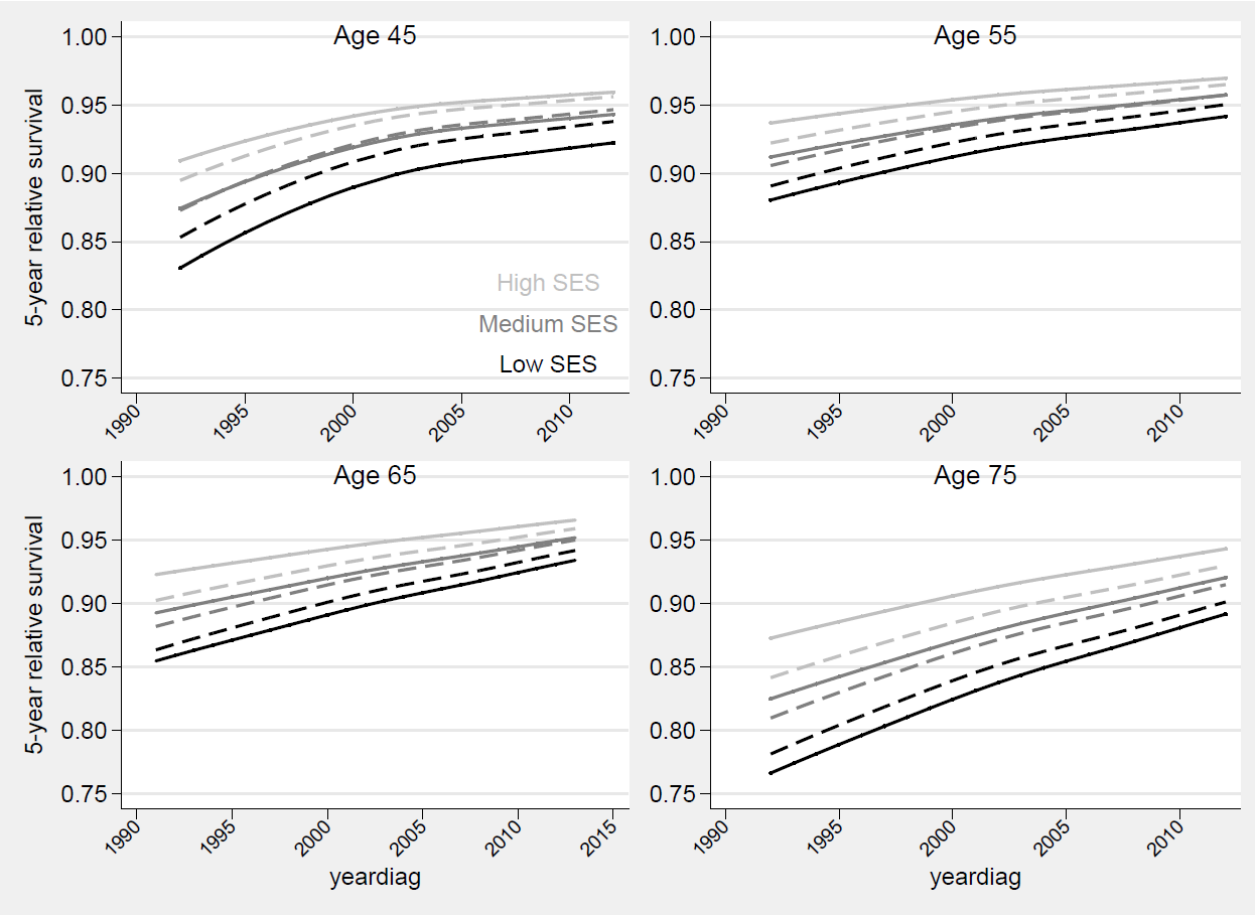

Figure 3: Five-year relative survival of female breast cancer patients diagnosed in the Stockholm and Gotland regions of Sweden using unadjusted life table rates (solid) and socioeconomic status (SES) -adjusted rates (dashed) from a flexible parametric survival model approach

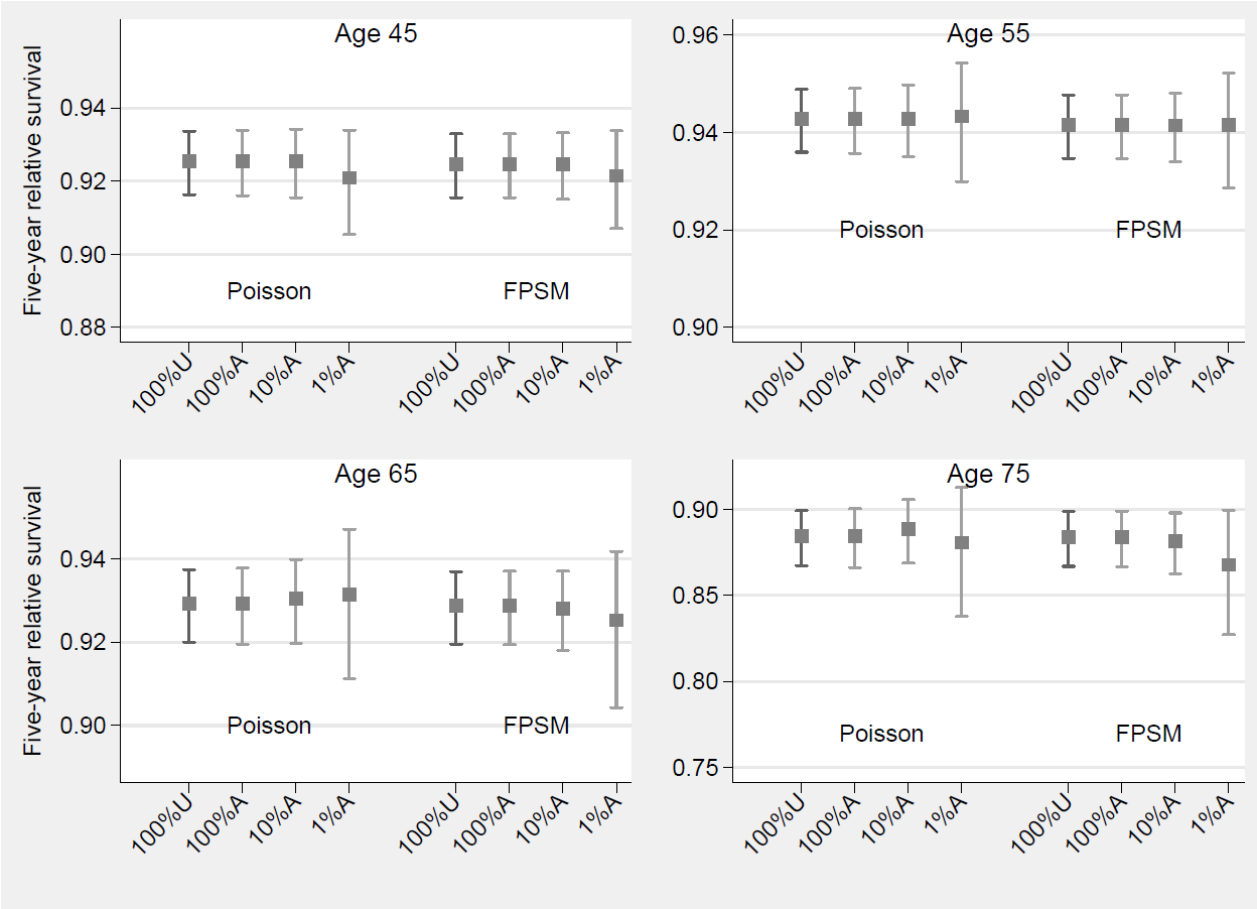

Figure 4: Five-year relative survival estimates and 95% confidence intervals (CIs) for high SES in 2000 from both Poisson models and flexible parametric survival models (FPSM) when using the full control dataset, n=133 361, and unadjusted CIs (100% U); the full control dataset and adjusted CIs (100% A); a random 10% sample, n=13 414; and 1% sample, n=1 394 of the control dataset and adjusted CIs (10% A and 1% A, respectively).

Tables

Table 1. Descriptive Statistics of Breast Cancer Patients and Controls From the Stockholm and Gotland Regions From BcBaSe Used in Analyses

	Breast cancer cases, n (%)	**Controls**, n (%)
	N= 26 913	N=133 361
Age; mean (SD)	61.3 (12.4)	61.3 (12.4)
Year		
1992-1994	2 892 (10.8)	14 345 (10.8)
1995-1999	5 760 (21.4)	28 510 (21.4)
2000-2004	6 512 (24.2)	32 174 (24.1)
2005-2009	7 032 (26.1)	34 891 (26.2)
2010-2012	4 717 (17.5)	23 441 (17.6)
SES group		
Low	6 873 (25.5)	37 792 (28.3)
Medium	10 491 (39.0)	53 225 (39.9)
High	9 549 (35.5)	42 344 (31.8)

Table 2. SES-Specific Standardized Mortality Ratio and 95% Confidence Intervals (CI) for Patients Diagnosed With Stage I Breast Cancer Between 2000 and 2010 Recorded in the Stockholm and Gotland Regions of BCBaSe Using SES-Adjusted, Poisson and Flexible Parametric Survival Model (FPSM) Approaches, and Unadjusted Expected Mortality Rate as the Reference

	Low SES		**Medium SES**		**High SES**	
	SMR	95% CI	SMR	95% CI	SMR	95% CI
Unadjusted	1.44	1.31,1.57	1.27	1.15,1.39	1.51	1.34,1.70
Poisson-adjusted	1.35	1.23,1.48	1.43	1.30,1.57	2.22	1.97,2.49
FPSM-adjusted	1.36	1.24,1.49	1.44	1.31,1.58	2.27	2.01,2.54