Time Trend Partitioning of Breast Cancer Prevalence and Mortality.

I. Akushevich[1], A.Yashkin[1], J. Kravchenko[2], B.Durgin[1], R.Greenup[2], S.Hwang[2], K.Land[1], A.I.Yashin[1]

[1]Biodemography of Aging Research Unit, Social Science Research Institute, Duke University, Durham, NC
[2]Department of Surgery, Duke University Medical Center, Durham, NC

Abstract
The time trend of female breast cancer (FBC) prevalence and mortality is the result of three competing processes: changes in the incidence rate, stage-specific survival, and ascertainment at early stages. Our new partitioning approach allows for the evaluation of the relative contributions of each of these components. Applying this approach to SEER data, we found that from 1993 the increase in prevalence was due to increased incidence (explains approximately 50% of total trend), improved survival (30-35%) and increased ascertainment at early stages (15-20%). The increase in FBC mortality up to 2002 was primarily due to the effect of increased incidence. After 2002 the trend goes to zero and changes sign (i.e., mortality decreases) due to the increased impact of survival. Partitioning approaches provide quantitative information on factors contributing to disease trends, a clear interpretation of the obtained results, and form the basis for both predictive inference and formal forecasting models.

Introduction
Breast cancer (BC) remains the most common cancer diagnosis among women in the United States[1] accounting for 29% of all new cancer cases in 2012. While advances in early detection and treatment have continued to improve breast cancer-specific mortality since 1990, it remains the second leading cause of cancer-related death for women in this country, accounting for 14% of all cancer deaths in 2015[1-4]. Advanced age remains the single most significant risk factor for breast cancer development, leading to a disproportionate number of breast cancer cases diagnosed in older women[5]. The probability of developing breast cancer markedly increases with age: from 1 in 203 at less than 39 years of age, to 1 in 28 at ages 60-69, to 1 in 15 at ages older than 70[1]. In 2011, approximately 78% of new breast cancer cases and 87% of breast cancer deaths occurred among women age 50 years and older, including over 22,660 deaths in the 65+[2] age group.

In this paper we focus on identifying historical and current gaps in breast cancer screening and treatment that have led to suboptimal survival in older women. Time trends in disease prevalence and especially mortality after BC diagnosis contain important information on periods of success and failure in breast cancer care and prevention. Epidemiologic trends of prevalence and BC-specific mortality represent highly reliable information directly related to the properties of BC care at the national level. Although BC is a heterogeneous disease with respect to patient and tumor characteristics, existing datasets have sufficient power to estimate time trends of prevalence and mortality in specific homogeneous groups. However, even within a relatively homogenous group, the time trend in associated prevalence/mortality is itself heterogeneous due to differences in the relative impact of the components that contribute to changes in prevalence and mortality over time. In the case of BC, the most influential components that can be reliably estimated with the available data are BC incidence, BC-stage at time of incident diagnosis, and BC-specific relative survival. The novel partitioning analysis used in this study creates reliable quantitative estimates of the relative contributions of the above components to the total trends in prevalence/mortality. Higher contributions of a specific component detected in a given patient stratum allow us to identify the *weak and modifiable links* in BC diagnosis or treatment. The cancer-specific partitioning approach to be used in this study is a further development a more general powerful statistical approach developed by our research team[6,7] to analyze the prevalence and mortality trends of diabetes mellitus using Medicare data.

Data
The SEER dataset (1973-2013) used in this study collects data on all cancer cases reported in SEER Registries covering about 26% of the U.S. It provides detailed information on tumor characteristics (stage, behavior, histology, grade, etc.) and patient survival as well as population counts in one-year groups for the calculation of age- and time-specific incidence rates. Prevalence of female breast cancer calculated using the partitioning approach developed in this paper are compared to the estimates given by i) SEER*Stat tool that produces cancer statistics (including cancer prevalence in population), and ii) SEER-Medicare data using approaches described in[8]. The calculated mortality is compared to estimates obtained using the Multiple Cause-of-Death (MCD) dataset (1968-2013) that collects information from death certificates for all cases of death in the U.S. providing both underlying and secondary causes.

Theory of partitioning analysis for breast cancer.

The partitioning approach developed in this study is based on an explicit representation of prevalence at ages 65+ with no simplifying assumptions. The resulting formulae for age-specific and age-adjusted prevalence for ages 65 and above are:

$$\begin{aligned} P(x,y) &= \int_0^{x-\bar{x}_0} I(\tau, y_d)\Big(\sum\nolimits_s \pi_s(\tau, y_d) S_d^r(x_\tau, \tau, y_d)\Big) dx_\tau \\ P(y) &= \int_{x_0}^{\infty}\Big(\int_0^{x-\bar{x}_0} I(\tau, y_d)\Big(\sum\nolimits_s \pi_s(\tau, y_d) S_d^r(x_\tau, \tau, y_d)\Big) dx_\tau\Big) p(x) dx \end{aligned} \quad (1)$$

where $P(x,y)$ and $P(y)$ are the age-specific and age-adjusted prevalence of a disease at age x and time y; $p(x)$ is the density of age distribution in a standard year, τ and $y_d = y - x + \tau$ are age and time of BC diagnosis, $x_\tau = x - \tau$ is the survival time. The index s runs over all stages at diagnosis. Three functions in the right hand side of eqn. (1) are i) $I(\tau, y_d)$, the incidence rate at age τ and year y_d, ii) $\pi_s(\tau, y_d)$, the probability that stage of diagnosis is s, and iii) $S_d^r(x_\tau, \tau, y_d)$, the relative survival function of individuals diagnosed at age x and year y_d with $x_\tau = x - \tau$ being survival time.

Similarly, the resulting formulae for age-specific and age-adjusted prevalence for ages 65 and above are:

$$\begin{aligned} M(x,y) &= \mu(x,y)P(x,y) + \int_{\bar{x}_0}^{x} I(\tau, y_d)\Big(\sum\nolimits_s \pi_s(\tau, y_d)\, f_d^r(x-\tau, \tau, y_d)\Big) d\tau \\ M(y) &= \int_{x_0}^{\infty}\Big(\mu(x,y)P(x,y) + \int_0^{x-\bar{x}_0} I(\tau, y_d)\Big(\sum\nolimits_s \pi_s(\tau, y_d)\, f_d^r(x_\tau, \tau, y_d)\Big) dx_\tau\Big) p(x) dx \end{aligned} \quad (2)$$

The formulae (1) and (2) are originated from the initial idea that probability of being prevalent at age x requires being incident at an earlier age τ $\tau,\ \tau \le x$ and having survived longer than $x - \tau$. Similarly, for mortality (we consider incidence-based mortality, i.e., mortality after disease onset) the probability of dying in the age interval $(x, x+dx)$ requires having death in the interval $(x, x+dx)$ and either being prevalent at the boundary point for this cohort or being incident at an earlier age $x - \tau$. Integration over all ages of diagnosis (τ) results in the formula for age-specific rates presented in eqn. (1). For mortality these transformations result in i) using the density f_d^r of the relative survival function and ii) occurring a separate term containing disease prevalence and mortality of the general population. Integration of age-specific prevalence and mortality over all ages results in the age-adjusted formula for the prevalence observed in a specific year y.

The mortality model (2) includes two terms: $M(y) = R_p + R_{is}$, where R_p represents the effect of mortality in the general population for individuals with breast cancer and R_{is} represents the effect of trends in cancer incidence, stage ascertainment, and cancer survival relative to general population survival. If relative survival is close to 1, then R_{is} would be small and $M(y) \approx R_p$, i.e., individuals with the disease would have mortality as in the general population. The situation is opposite for breast cancer and we expect $R_p \ll R_{is}$.

The time trends in $P(y)$ and $M(y)$ are determined by the trends in the incidence rate, stage-specific survival, and in ascertainment at early stages as well as mortality in the general population in the case of $M(y)$. Formally the time trend is defined as the derivative of $P(y)$ and $M(y)$ with respect to y or as the annual percent change. The explicit calculation of this derivative results in:

$$\begin{aligned} \frac{P'(y)}{P(y)} = \frac{1}{P(y)} \int_{x_0}^{\infty}\int_0^{x-\bar{x}_0} \sum\nolimits_s \Big(& I'(\tau, y_d)\pi_s(\tau, y_d) S_s^r(x_\tau, \tau, y_d) \\ & + I(\tau, y_d)\pi_s'(\tau, y_d) S_s^r(x_\tau, \tau, y_d) + I(\tau, y_d)\pi_s(\tau, y_d) S_s^{r\prime}(x_\tau, \tau, y_d)\Big) p(x) dx_\tau dx \end{aligned} \quad (3)$$

Thus, the time trend of disease prevalence is determined through three terms within the brackets of eqn. (3) which correspond to the contributions of incidence, ascertainment at early stages, and patient survival. The eqn. (3) represents the partitioning of time trends in the prevalence:

$$P'(y) / P(y) = T_{inc}(y) + T_\pi(y) + T_S(y) \quad (4)$$

Time trend partitioning for mortality is obtained similar excepting that two additional terms (disease prevalence and mortality in the general population) also contribute. The result is:

$$M'(y)/M(y) = \hat{T}_P(y) + \hat{T}_\mu(y) + \hat{T}_{inc}(y) + \hat{T}_\pi(y) + \hat{T}_S(y) \quad (5)$$

Explicit expressions for $T_{inc,\pi,S}(y)$ and $\hat{T}_{P,\mu,inc,\pi,S}(y)$ are obtained given specific models of age- and time-dependence if model components are developed and estimated.

Partitioning analysis – Estimation.

The quantities of interest (i.e., $T_{...}(y)$ and $\hat{T}_{...}(y)$) are expressed in terms of derivatives of the respective functions with respect to time. In our approach, we use explicit analytic parameterizations for all functions for which derivatives are needed. Therefore, we calculate the derivatives analytically. This allows us to avoid dealing with possible numeric instabilities occurring when derivatives are evaluated numerically. For any function we use a two-stage approach: i) we parameterize the function and fit data for each year of diagnosis, and ii) use B-splines to model relationships between year-specific model parameters and evaluate y-dependences of the function. B-splines allow explicit calculation of derivatives without requiring additional simplifying assumptions. The analysis involves the design and estimation of four models: i) for the incidence rate, ii) for the probability of relative survival after cancer diagnosis, iii) for frequencies of stage at onset, and iv) for mortality in the general population. The distribution of age and time after onset was modeled using the Armitage-Doll model with additional individual predisposition modeled by gamma or inverse Gaussian distributions[9] (for incidence), the Weibull model for time after disease onset (for survival) and the Gompertz model (for mortality). We proved that this model is successful; time trends were modeled by estimating model parameters for each year and apply B-splines to fit the time patterns of obtained parameters. Prevalence of breast cancer calculated using the partitioning approach developed in this paper compare favorably to the estimates given by i) SEER*Stat tool that produces cancer statistics (including cancer prevalence in population), and ii) SEER-Medicare data using approaches described in Akushevich et al.[8].

Results.

Our incidence model[9-11] was shown to fit the empirical age-specific incidence rates (Figure 1) and empirical age-adjusted incidence (Figure 2) with high precision. Age-adjusted breast cancer incidence reached a peak on or about 2000 and has been declining since. Changes in the relative frequencies of cancer stage (local, regional, distant) at diagnosis (Figure 3) are almost flat in the study period, and the model represents the patterns well. However the general tendency is that female breast cancer is being diagnosed at earlier stages. The relative survival is parameterized using the Weibul model. This choice is justified by the linear nature of the empiric rates presented in $\log\left(-\log S(x_\tau)\right)$ vs. $\log(x_\tau)$ (Figure 4).

The left plot of Figure 5 presents age-adjusted breast cancer prevalence obtained using our model (1) in comparison with estimates obtained using SEER*Stat for 18 and 23 years of look-back period as well as that observed in SEER-Medicare for a 6-month look-back period. The agreement between empirical data and the model is good: the shape of the model mirrors empirical patterns, and differences in magnitude are minor. The results of our mortality model are presented in the right plot of Figure 6. Our model (2) shows good agreement with rates derived from the SEER*Stat that provide incidence-based mortality obtained from SEER data. As expected, both contributions to mortality (given by eq. (2)) are similar the term with new cases dominated before 1994 ($R_p < R_{is}$), and the term describing mortality of prevalent individuals is higher starting from 1994 ($R_p > R_{is}$). The results show that both prevalence and incidence-based mortality has been increasing during the study period.

Partitioning of disease prevalence (Figure 6, left plot) and incidence–based mortality (Figure 6, right plot) contain three (trends of incidence, stage ascertainment, and relative survival) and five (trends of incidence, stage ascertainment, relative survival, and prevalence/mortality in the general population) contributing components respectively. The increase in breast cancer prevalence for females is due to an increased incidence (explains approximately 50% of total trend) and improved survival (30-35% of total trend). The remaining 15-20% are explained by increased ascertainment at early stages. These percentages are valid for 1995. Before 1995 (especially before 1985) the increased prevalence was (up to 80%) due to trends in incidence and after 1985 trends in survival became contributing and after 2004 became dominant. Similarly, the increase in mortality for females up to 2002 is due to the effect of increased incidence. After 2002 the trend goes to zero and changes sign (i.e., mortality decreases). This happened due to the increased impact of survival after 2002. For males, the effect of increased survival is compensated by a rapidly declining incidence rate resulting in an overall decrease in prevalence over time. Stage ascertainment plays an adverse role in

prevalence for males with maximal contribution in 2005 (it compensates about 30% of the contribution of decreased incidence). Declining incidence is the main factor decreasing BC mortality for females, though improved survival also pushes mortality down starting from 2002. For example, in 2005 the contribution of relative survival and incidence decreased mortality were about 30% and 60%. The remaining 10% are due to decreased mortality in the general population.

Discussion and Conclusion.

The approach developed in this paper provides high precision estimates of the time-trends of female breast cancer prevalence and mortality and explains them in terms of trends of breast cancer incidence rate, stage-specific relative survival, and ascertainment at early stages. All empirically estimated age- and time-patterns of constituent components are in agreement to those known in literature. We found that survival from breast cancer increases its prevalence for both genders. This demonstrates success of health care in breast cancer treatment. Female breast cancer incidence increased over the entire period though the effect became smaller after 2000. Partitioning of prevalence and mortality shows that the contribution of stage ascertainment has a beneficial effect with maximum in 2005, although this effect is small when compared to incidence and survival. Since stage at diagnosis is one of the primary determinants of breast-cancer mortality this indicates the need for further improved screening methods to improve early stage ascertainment.

In conclusion, partitioning approaches such as the one developed in this study provide quantitative information on factors contributing to disease trends, a clear interpretation of the obtained results and can form the basis for both predictive inference and formal forecasting models. Specifically, partitioning reveals the mechanisms of the interrelations between cancer incidence, prevalence, and survival—an important step in assessing the potential burden posed by a disease, potential problem areas in need of attention, as well as how these important epidemiological characteristics respond to implemented health interventions over time. In the case of breast cancer, the main driving forces underlying the observed trends in prevalence and mortality were: i) increasing incidence staring from 1983 and deceleration in this increase since 1990 that probably reflects changes in the hormone replacement therapy; ii) improved stage-specific survival also starting from 1983. Incidence is the primary determinant of female prevalence trends. Based on the results of our analysis we infer that further improvement in breast cancer mortality can be achieved by improvement of early stage ascertainment (that became non-beneficial last years) on par with breast cancer incidence that continue having disadvantageous trends.

References.

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. *CA Cancer J Clin.* 2012.
2. DeSantis C, Siegel R, Bandi P, Jemal A. Breast cancer statistics, 2011. *CA Cancer J Clin.* 2011.
3. DeSantis CE, Fedewa SA, Goding Sauer A, Kramer JL, Smith RA, Jemal A. Breast cancer statistics, 2015: Convergence of incidence rates between black and white women. *CA: a cancer journal for clinicians.* 2016;66(1):31-42.
4. SEER S. Stat Database: Incidence-SEER 9 Regs Research Data, Nov 2015 Sub (1973-2013)< Katrina/Rita Population Adjustment>− Linked To County Attributes− Total US, 1969-2014 Counties. 2015. *Surveillance Research Program, Surveillance Systems Branch.* 2014.
5. Giordano SH, Hortobagyi GN, Kau S-WC, Theriault RL, Bondy ML. Breast cancer treatment guidelines in older women. *Journal of Clinical Oncology.* 2005;23(4):783-791.
6. Akushevich I, Yashkin A, Kravchenko J, et al. Theory of partitioning of disease prevalence and mortality in observational data. *Theoretical Population Biology.* 2017;114:117-127.
7. Akushevich I, Yashkin AP, Kravchenko J, et al. Identifying the Causes of the Changes in the Prevalence Patterns of Diabetes in Older US Adults: A New Trend Partitioning Approach. *Journal of Diabetes and its Complications.* 2018;32(4):362-367.
8. Akushevich I, Kravchenko J, Arbeev KG, Ukraintseva SV, Land KC, Yashin AI. Health Effects and Medicare Trajectories: Population-Based Analysis of Morbidity and Mortality Patterns. *Biodemography of Aging*: Springer; 2016:47-93.
9. Kravchenko J, Akushevich I, Abernethy AP, Lyerly HK. Evaluating the number of stages in development of squamous cell and adenocarcinomas across cancer sites using human population-based cancer modeling. *PloS one.* 2012;7(5):e37430.

10. Kravchenko J, Akushevich I, Seewaldt VL, Abernethy AP, Lyerly HK. Breast cancer as heterogeneous disease: contributing factors and carcinogenesis mechanisms. *Breast cancer research and treatment.* 2011;128(2):483-493.
11. Il'yasova D, Siamakpour-Reihani S, Akushevich I, Akushevich L, Spector N, Schildkraut J. What can we learn from the age-and race/ethnicity-specific rates of inflammatory breast carcinoma? *Breast cancer research and treatment.* 2011;130(2):691.

Figures

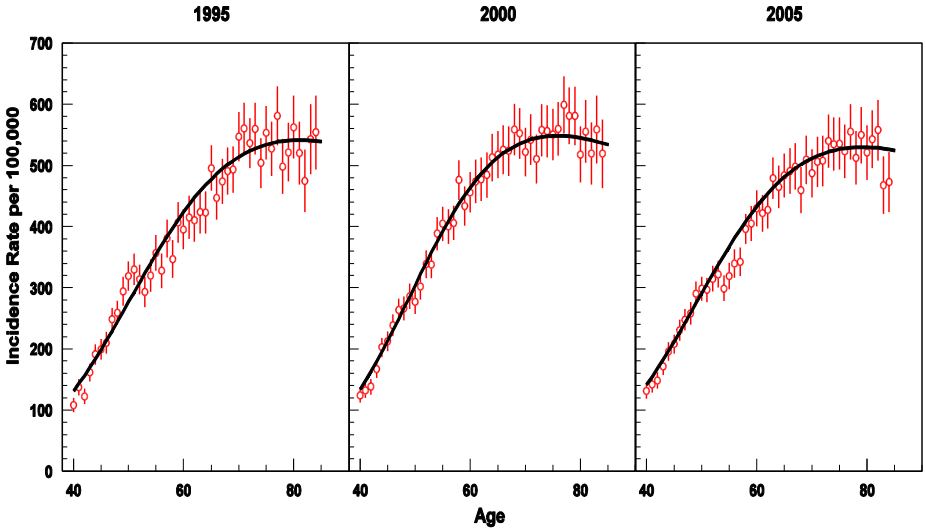

Figure 1. Age-patterns of breast cancer incidence rates: empiric estimates using the SEER data (open dots with errors bars) and the estimated Armitage-Doll model with random predisposition.

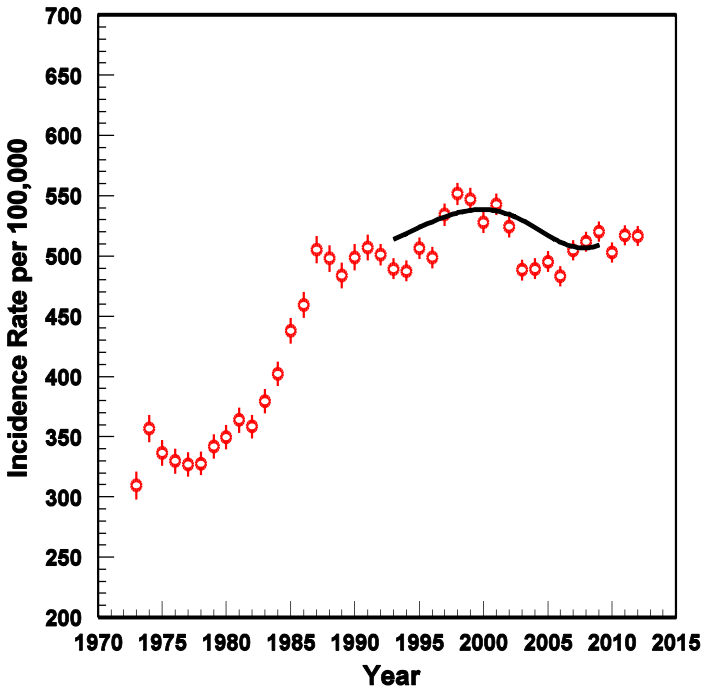

Figure 2. Age-adjusted (65+) incidence rate of breast cancer incidence: empiric estimates using the SEER data (open dots with errors bars) and the B-spline model for calendar year dependence. The year dependence of model parameters using B-splines were first calculated resulting in a model for age pattern for any year and then age adjusted rates were calculated

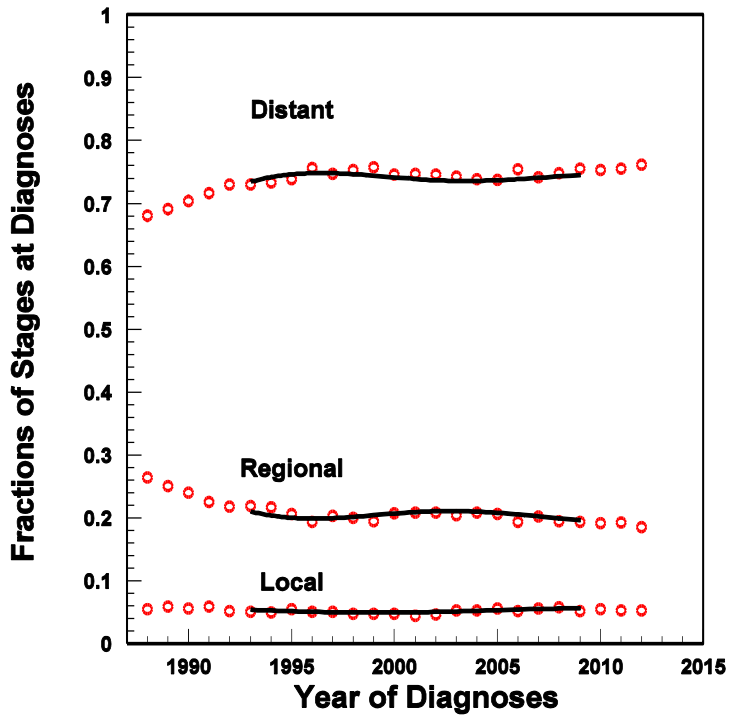

Figure 3. Empiric estimates (open dots with error bars) and B-spline model (curves) of the year dependence of frequencies of diagnoses at a specific stage (f_{si})

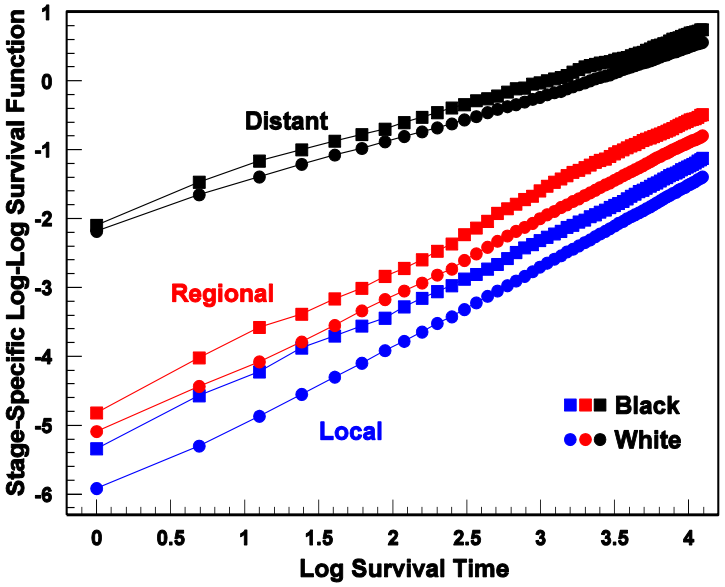

Figure 4. Stage and race-specific survival functions (in the $\log(-(\log(S(t)))$ vs. $\log(t)$ format) for individuals diagnosed in 1995-2007

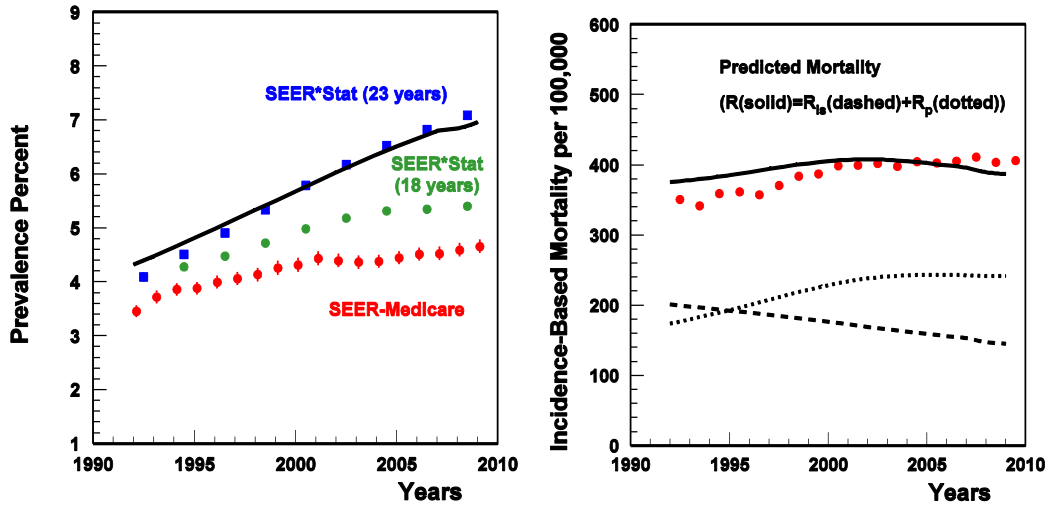

Figure 5. Breast cancer prevalence (left panel): predicted using our model (solid black line), calculated using SEER*Stat for 18 (green points) and 23 (blue points) years of look-back period, and observed in SEER-Medicare for 6-month look-back period (red points). Breast cancer mortality (right panel): predicted using our model (solid black line) and the two terms R_p (dotted line) and R_{is} (dashed line), and calculated incidence-based mortality using SEER*Stat (red points).

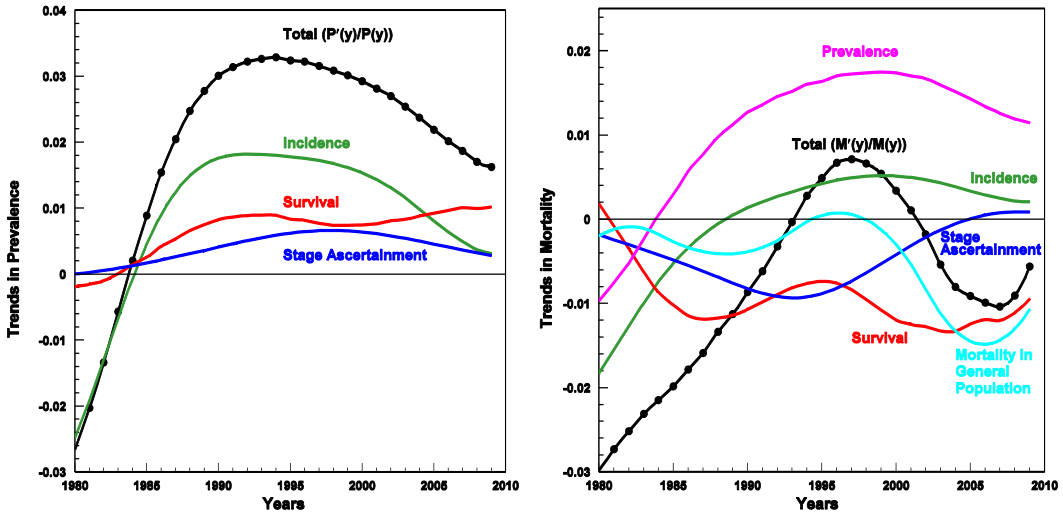

Figure 6. Partitioning of the prevalence trends (left panel) according its contributions: $P'(y)/P(y) = T_{inc}(y) + T_{st}(y) + T_{surv}(y)$. Partitioning of the mortality trends (right panel) according its contributions: $M'(y)/M(y) = \hat{T}_P(y) + \hat{T}_\mu(y) + \hat{T}_{inc}(y) + \hat{T}_\pi(y) + \hat{T}_S(y)$