BMC Cancer

Research article

Global and regional estimates of cancer mortality and incidence by site: I. Application of regional cancer survival model to estimate cancer mortality distribution by site

Colin D Mathers*[1], Kenji Shibuya[1], Cynthia Boschi-Pinto[2], Alan D Lopez[1] and Christopher JL Murray[3]

Address: [1]Global Programme on Evidence for Health Policy, World Health Organization, Geneva, Switzerland, [2]Family and Community Health/ Child and Adolescent Health and Development, World Health Organization, Geneva, Switzerland and [3]Executive Director, Evidence and Information for Policy, World Health Organization, Geneva, Switzerland

Email: Colin D Mathers* - mathersc@who.ch; Kenji Shibuya - shibuyak@who.int; Cynthia Boschi-Pinto - pintoc@who.ch; Alan D Lopez - lopeza@who.ch; Christopher JL Murray - murrayc@who.ch

* Corresponding author

Abstract

Background: The Global Burden of Disease 2000 (GBD 2000) study starts from an analysis of the overall mortality envelope in order to ensure that the cause-specific estimates add to the total all cause mortality by age and sex. For regions where information on the distribution of cancer deaths is not available, a site-specific survival model was developed to estimate the distribution of cancer deaths by site.

Methods: An age-period-cohort model of cancer survival was developed based on data from the Surveillance, Epidemiology, and End Results (SEER). The model was further adjusted for the level of economic development in each region. Combined with the available incidence data, cancer death distributions were estimated and the model estimates were validated against vital registration data from regions other than the United States.

Results: Comparison with cancer mortality distribution from vital registration confirmed the validity of this approach. The model also yielded the cancer mortality distribution which is consistent with the estimates based on regional cancer registries. There was a significant variation in relative interval survival across regions, in particular for cancers of bladder, breast, melanoma of the skin, prostate and haematological malignancies. Moderate variations were observed among cancers of colon, rectum, and uterus. Cancers with very poor prognosis such as liver, lung, and pancreas cancers showed very small variations across the regions.

Conclusions: The survival model presented here offers a new approach to the calculation of the distribution of deaths for areas where mortality data are either scarce or unavailable.

Background

As a part of the Global Burden of Disease 2000 (GBD 2000) project, the present study aims at estimating the total global and regional cancer mortality and incidence based on its detailed analysis of all-cause levels and cause of death distributions for 191 Member States of the World Health Organization (WHO) [1]. GBD 2000 requires age- and sex- specific incidence, duration and mortality as a

minimum input to estimate burden of each disease sequela by a composite measure of mortality and morbidity (i.e., disability-adjusted life years: DALYs).

Attempts have been made to quantify the global burden of cancer, and estimate site-specific cancer mortality and morbidity [2–6]. Such studies have repeatedly suggested that incidence and mortality from cancer are continuously increasing in many parts of the world. Despite an increasing trend of cancer incidence and mortality, data on survival and prognosis of incident cases from population-based cancer registries are limited in majority of developing countries. One of the most credible sources of information is available from the International Agency for Research on Cancer (IARC), which has been coordinating and implementing the cancer registries in such regions [7].

While vital registration of causes of death and national cancer registries are perhaps the best source of data on cancer mortality, mortality data are still scarce, poor or even unavailable for some regions of the world. Innovative methods will thus continue to be needed to exploit available data. Estimating mortality from morbidity and, especially, morbidity from mortality was a common practice in the 70's and 80's [8,9]. More recently, the continuos effort made by IARC has lead to the Globocan 2000 estimates which has also used information on incidence and survival to estimate cancer death for the year 2000 from various sources including cancer registries [2,6]. Still others have made use of vital statistics and cancer incidence data to predict the number of new cancer cases and deaths for the US in the subsequent year [10].

On the basis of available published information on age-, sex-, and site-specific cancer incidence and survival, we developed an algorithm to estimate region-specific overall cancer mortality, and site-specific survival, death distributions and incidence for the year 2000. This paper presents the first of the two consecutive reports which present the detailed methods and results of GBD 2000 estimates for mortality and incidence of cancer by site.

The particular feature of the GBD 2000 study is that the number of deaths by age and sex in each region provides an essential envelope which constrains individual disease and injury estimates of deaths and that competing claims for the magnitude of deaths from various causes must be reconciled within this envelope [1]. Given the regional cancer mortality envelope by age and sex, the estimates of site-specific distributions of cancer mortality are necessary to disaggregate the estimated total cancer deaths by age and sex for each region.

For geographic disaggregation of the GBD 2000, the six WHO regions of the world have been further divided into 14 sub regions, based on levels of child (under five years) and adult (15–59 years) mortality for WHO Member States [1]. Five mortality strata were defined in terms of quintiles of the distribution of child and adult mortality (both sexes combined). Adult mortality was regressed on child mortality and the regression line used to divide countries with high child mortality into high adult mortality (stratum D) and very high adult mortality (stratum E). Stratum E includes the countries in sub-Saharan Africa where HIV/AIDS has had a very substantial impact.

When these mortality strata are applied to the six WHO regions, they produce 14 mortality subregions. For the purposes of burden of disease epidemiological analyses, 2 of these regions were further subdivided: EurB into EurB1 and EurB2 – the latter including the central Asian states; and WprB into WprB1 (mainly China), WprB2 (South East Asian countries) and WprB3 (Pacific Islands). Additionally, some Member States have been reclassified into subregions with similar epidemiological/geographic/ethnic patterns in order to maximise the epidemiological homogeneity of the subregions for the purposes of epidemiological analysis. The resulting 17 epidemiological subregions are listed in Table 1.

The approaches to estimating mortality distributions were different depending on the availability and quality of data on detailed causes of death. Direct estimates of the site-specific distributions of cancer mortality were possible for the regions where established vital registration records with high coverage and coding practice based on the International Statistical Classification of Diseases and Related Health Problems (ICD) are available, including countries in the A sub regions (AmrA, EurA and WprA) and countries in AmrB, EurB1, EurB2 and EurC [1]. For the other regions of the world (AfrD, AfrE, AmrD, EmrB, EmrD, SearB, SearD, WprB1, WprB2 and WprB3), we developed a site-specific model for relative interval survival adjusted for each region and applied it to the regional incidence estimated to calculate the mortality distribution by site for the year 2000. This model can also be used to estimate survival at different ages and average duration of cancer by site. In this paper, we present a detailed model as a key input to estimate the distribution of cancer deaths by site model for the regions where few data are available.

Material and Methods
Data sources
Relative interval survival based on the US data
The primary data sources used to develop the cancer survival model were the National Cancer Institute Surveillance, Epidemiology, and End Results (SEER) statistical program (SEER*Stat version 4.2). The SEER program is

Table 1: Global Burden of Disease 2000 (GBD 2000) project: regions and sub regions

WHO region	Mortality stratum	Sub region	WHO Member States
AFRO	D	AfrD	Algeria, Angola, Benin, Burkina Faso, Cameroon, Cape Verde, Chad, Comoros, Equatorial Guinea, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Liberia, Madagascar, Mali, Mauritania, Mauritius, Niger, Nigeria, Sao Tome And Principe, Senegal, Seychelles, Sierra Leone, Togo, Djibouti, Somalia, Sudan
AFRO	E	AfrE	Botswana, Burundi, Central African Republic, Congo, Côte d'Ivoire, Democratic Republic Of The Congo, Eritrea, Ethiopia, Kenya, Lesotho, Malawi, Mozambique, Namibia, Rwanda, South Africa, Swaziland, Uganda, United Republic of Tanzania, Zambia, Zimbabwe
AMRO	A	AmrA	Canada, United States Of America
AMRO	B	AmrB	Antigua And Barbuda, Argentina, Bahamas, Barbados, Belize, Brazil, Chile, Colombia, Costa Rica, Dominica, Dominican Republic, El Salvador, Grenada, Guyana, Honduras, Jamaica, Mexico, Panama, Paraguay, Saint Kitts And Nevis, Saint Lucia, Saint Vincent And The Grenadines, Suriname, Trinidad And Tobago, Uruguay, Venezuela Cuba
AMRO	D	AmrD	Bolivia, Ecuador, Guatemala, Haiti, Nicaragua, Peru
EMRO	B	EmrB	Bahrain, Cyprus, Iran (Islamic Republic Of), Jordan, Kuwait, Lebanon, Libyan Arab Jamahiriya, Oman, Qatar, Saudi Arabia, Syrian Arab Republic, Tunisia, United Arab Emirates
EMRO	D	EmrD	Egypt, Iraq, Morocco, Yemen
EURO	A	EurA	Andorra, Austria, Belgium, Croatia, Czech Republic, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Israel, Italy, Luxembourg, Malta, Monaco, Netherlands, Norway, Portugal, San Marino, Slovenia, Spain, Sweden, Switzerland, United Kingdom
EURO	B1	EurB1	Albania, Bosnia And Herzegovina, Bulgaria, Georgia, Poland, Romania, Slovakia, The Former Yugoslav Republic Of Macedonia, Turkey, Yugoslavia
EURO	B2	EurB2	Armenia, Azerbaijan, Kyrgyzstan, Tajikistan, Turkmenistan, Uzbekistan
EURO	C	EurC	Republic of Moldova, Russian Federation, Ukraine
SEARO	B	SearB	Indonesia, Sri Lanka, Thailand Malaysia, Philippines Brunei Darussalam, Singapore
SEARO	D	SearD	Bangladesh, Bhutan, India, Maldives, Nepal Afghanistan, Pakistan
WPRO	A	WprA	Australia, Japan, New Zealand
WPRO	B1	WprB1	China, Mongolia, Republic Of Korea DPR Korea
WPRO	B2	WprB2	Cambodia, Lao People's Democratic Republic, Viet Nam Myanmar
WPRO	B3	WprB3	Cook Islands, Fiji, Kiribati, Marshall Islands, Micronesia (Federated States Of), Nauru, Niue, Palau, Papua New Guinea, Samoa, Solomon Islands, Tonga, Tuvalu, Vanuatu

considered as the standard for quality among cancer registries around the world, being the most authoritative source of information on cancer incidence and survival in the United States [11], *RIS* was directly obtained from the SEER database (1973–1997 Public-Use Data) within SEER*Stat for every age group, sex, and cancer site. Cancer sites for which survival was calculated were: mouth and pharynx (ICD-10 C00-C14), oesophagus (C15), stomach (C16), colon and rectum (C18-C21), liver (C22), pancreas (C25), trachea, bronchus and lung (C33-C34), melanoma of the skin (C43), female breast (C50), cervix uteri (C53), corpus uteri (C54-55), ovary (C56), prostate (C61), bladder (C67), lymphomas and multiple myeloma (C81-C90, C96), leukaemia (C91-C95), and other malignant neoplasms (balance of ICD-10 C00-C97).

Incidence data

We initially used the Globocan 2000 estimates of the international Agency for Research on Cancer (IARC) to apply the survival model for a region [6], assuming that incidence rates to be constant over the years. We then estimated the region-specific number of new cases for 1986 to 2000, by applying these age-specific incidence rates to the annual population. We carefully examined the methods used to estimate country-specific incidence data in Globocan 2000, to ensure that for all the regions where we required incidence estimates, the Globocan estimates were based on cancer registry incidence data, and not modelled from mortality data using assumptions about survival (which would then result in circularity in our mortality estimation process for regions without good mortality data by cancer site).

Table 2: Estimated relative probability of death after 1 year (1 - *RIS*₁)

Site	Males	Females
Mouth and pharynx	0.202	0.180
Oesophagus	0.660	0.635
Stomach	0.574	0.558
Colon and rectum	0.207	0.223
Trachea, bronchus and lung	0.831	0.769
Pancreas	0.828	0.816
Lung	0.631	0.568
Melanoma of the skin	0.052	0.029
Breast		0.039
Cervix		0.126
Uterus		0.073
Ovary		0.289
Prostate	0.037	
Bladder	0.091	0.153
Non-Hodgkin lymphoma	0.297	0.253
Hodgkin lymphoma	0.108	0.089
Multiple myeloma	0.286	0.297
Leukemia	0.360	0.376
Others	0.348	0.386

Globocan 2000 estimates of cancer incidence by site for countries differ from those required for the GBD 2000 in two major respects: 1) Globocan 2000 estimates include Kaposi's sarcoma and non-Hodgkin lymphomas (NHL) caused by HIV/AIDS. The GBD 2000 includes these cases among AIDS sequela and their burden is included with the HIV/AIDS burden [12–14] and 2) Globocan 2000 estimates include cancers of unknown primary with cancers of other specified sites. The GBD 2000 attributes these ill-defined cancers back to specific sites as described above. Accordingly Globocan 2000 incidence estimates by age, sex, site and country were adjusted for these differences. Firstly, unpublished data on the incidence of Kaposi's sarcoma for countries in Africa were provided by IARC and used to adjust incidence of other cancers to remove incidence of Kaposi's sarcoma. Secondly, relative risks of HNL from HIV [15–18] were estimated and, together with the UNAIDS prevalence estimates of HIV in each country of the African region, NHL attributable to HIV was also removed. Thirdly, incidence estimates for cancers of unknown primary site were redistributed among specific sites using the GBD 2000 algorithm [1]. The proportion of the others category (balance of all but skin cancers) in the Globocan 2000 corresponding to unknown primary sites was estimated from published data on the distribution of cancer incidence by site which included unknown primary as a specific category [19–27].

After adjusting the Globocan 2000 incidence estimates for each country as described above, these estimates were summed for the countries in each GBD 2000 region, re-

sulting in estimated incidence distributions by site, age and sex for each region. Finally, the GBD 2000 uses the latest population estimates for the Member States of the World Health Organization prepared by the United Nations Population Division [28]. In order to obtain incidence from 1986 to 2000, we estimated the age-specific population by sex for each of these years, using growth rates also from the United Nation's data.

Multiplicative model of relative interval survival
In order to estimate cancer death distribution for regions where little cancer mortality data is available (AfrD, AfrE, AmrD, EmrB, EmrD, SearB, SearD, WprB1, WprB2 and WprB3 sub regions), we developed an age-period-cohort multiplicative model for the relative interval survival (RIS) for each site. To incorporate all three time dimensions, we have taken into account the relative survival for every 5-year age group from 0 up to 85+ years of age, and for calendar year for 15 years (1981 to 1995), and for time since cancer diagnosis (survival time for cohorts) from 1-up to 15-year survival. After obtaining the time-specific survival data, we have then further indexed all the age, time, and calendar year survival information to the first year interval survival for each sex, and cancer site. The first year of survival was chosen because, for most if not all cancer sites, it is the most critical year concerning cancer survival experience. After the first year of survival, the relative survival curve usually increases and then flattens smoothly. Indexing was done by dividing each of the time-specific RIS by the survival at 1-year interval.

Table 3: Estimated age parameters (A_α) by cancer site, age and sex

Age	Oral Male	Oral Female	Oesophagus Male	Oesophagus Female	Stomach Male	Stomach Female	Colon and rectum Male	Colon and rectum Female	Liver Male	Liver Female	Pancreas Male	Pancreas Female
0–4	0.870	1.093	0.957	0.944	0.864	0.814	1.169	0.288	0.330	0.388	0.824	0.408
5–9	0.429	0.222	0.957	0.944	0.864	0.814	1.169	0.288	0.430	0.600	0.824	0.408
10–14	0.216	0.258	0.957	0.944	0.864	0.814	1.169	0.288	0.527	0.315	0.824	0.613
15–19	0.185	0.200	0.957	0.944	0.864	0.814	1.169	0.688	0.225	0.585	0.824	0.204
20–24	0.337	0.099	0.957	0.944	0.864	0.895	1.169	0.467	0.580	0.638	0.824	0.204
25–29	0.710	0.177	0.957	0.944	0.864	0.888	1.169	0.820	0.818	0.689	0.824	0.531
30–34	0.913	0.314	0.957	0.944	0.864	0.685	1.169	0.590	0.949	0.773	0.824	0.497
35–39	0.761	0.436	0.957	0.888	0.864	0.811	0.966	0.700	0.960	0.752	0.893	0.712
40–44	0.818	0.526	0.957	0.985	0.864	0.746	0.765	0.739	1.033	0.780	0.907	0.745
45–49	0.883	0.631	0.965	0.977	0.931	0.791	0.787	0.687	0.974	0.819	0.927	0.902
50–54	0.908	0.894	0.969	1.036	0.899	0.851	0.784	0.671	0.980	0.969	0.936	0.922
55–59	0.936	0.939	0.948	0.996	0.877	0.816	0.797	0.739	1.010	0.962	0.977	0.952
60–64	1.017	0.884	0.967	0.959	0.940	0.863	0.812	0.749	1.023	0.998	0.986	0.970
65–69	1.048	1.086	1.029	0.949	0.972	0.923	0.861	0.857	1.010	1.025	1.015	0.989
70–74	1.145	1.103	1.038	0.975	0.990	0.982	0.979	0.898	1.051	1.034	1.016	1.020
75–79	1.141	1.304	1.063	0.969	1.082	1.055	1.134	1.069	1.032	1.082	1.016	1.033
80–84	1.220	1.354	1.096	1.066	1.169	1.121	1.368	1.275	1.070	1.126	1.061	1.058
85+	1.358	1.993	1.152	1.210	1.259	1.279	1.793	1.619	1.039	1.126	1.082	1.084

Age	Lung Male	Lung Female	Bladder Male	Bladder Female	Non-Hodgkin lymphoma Male	Non-Hodgkin lymphoma Female	Hodgkin lymphoma Male	Hodgkin lymphoma Female	Multiple myeloma Male	Multiple myeloma Female	Leukaemia Male	Leukaemia Female
0–4	0.711	0.849	0.107	0.141	0.824	0.823	0.476	0.000	1.000	1.000	0.356	0.354
5–9	0.711	0.849	0.107	0.141	0.716	0.502	0.304	0.972	1.000	1.000	0.280	0.224
10–14	0.711	0.849	0.107	0.141	0.791	0.664	0.226	0.493	1.000	1.000	0.483	0.525
15–19	0.711	0.849	0.107	0.141	0.850	0.859	0.159	0.191	1.000	1.000	0.703	0.684
20–24	0.711	0.849	0.107	0.141	0.953	0.767	0.303	0.204	1.000	1.000	0.865	0.815
25–29	0.711	0.849	0.107	0.159	1.261	0.673	0.280	0.161	0.431	0.318	0.883	0.936
30–34	0.926	0.849	0.182	0.393	1.381	0.677	0.507	0.259	0.670	0.619	0.814	0.796
35–39	0.939	0.849	0.157	0.533	1.320	0.565	0.492	0.422	0.485	0.540	0.866	0.872
40–44	0.922	0.890	0.344	0.459	1.107	0.521	0.794	0.346	0.591	0.622	0.702	0.801
45–49	0.909	0.891	0.443	0.452	0.872	0.526	1.145	0.744	0.556	0.525	0.770	0.882
50–54	0.920	0.891	0.459	0.405	0.715	0.576	1.262	1.084	0.700	0.725	0.780	0.830
55–59	0.932	0.918	0.550	0.522	0.713	0.630	1.401	1.683	0.684	0.669	0.811	0.829
60–64	0.954	0.937	0.756	0.637	0.809	0.760	2.194	2.317	0.800	0.785	0.899	0.880
65–69	0.986	0.979	0.751	0.764	0.860	0.877	2.581	3.020	0.898	0.762	0.991	0.959
70–74	1.039	1.038	1.069	0.878	1.003	1.058	3.753	3.912	1.139	0.942	1.192	1.068
75–79	1.090	1.038	1.305	1.199	1.251	1.428	4.263	4.686	1.231	1.196	1.404	1.185
80–84	1.169	1.038	1.740	1.520	1.531	1.654	4.979	4.303	1.535	1.451	1.501	1.337
85+	1.219	1.038	2.453	1.956	1.863	2.105	5.372	6.393	1.828	1.743	1.563	1.535

Age	Melanoma Male	Melanoma Female	Miscellaneous Male	Miscellaneous Female	Breast Female	Cervix Female	Uterus Female	Ovary Female	Prostate Male
0–4	1.000	3.573	0.514	0.430	0.874	0.574	0.322	0.247	0.548
5–9	1.000	1.000	0.444	0.405	0.874	0.574	0.322	0.247	0.548
10–14	0.324	0.706	0.395	0.299	0.874	0.574	0.322	0.247	0.548
15–19	0.959	0.937	0.377	0.222	0.874	0.574	0.322	0.202	0.548
20–24	0.670	0.531	0.279	0.147	0.874	0.425	0.322	0.241	0.548
25–29	0.764	0.455	0.229	0.154	0.679	0.329	0.553	0.173	0.548
30–34	0.564	0.493	0.261	0.207	0.733	0.406	0.382	0.236	0.548
35–39	0.902	0.542	0.375	0.270	0.651	0.457	0.353	0.336	0.548

Table 3: Estimated age parameters (A$_\alpha$) by cancer site, age and sex *(Continued)*

40–44	0.857	0.769	0.527	0.400	0.569	0.508	0.458	0.433	0.548
45–49	0.871	0.710	0.709	0.498	0.564	0.739	0.411	0.496	0.548
50–54	0.815	1.073	0.850	0.620	0.790	0.994	0.473	0.614	0.548
55–59	1.110	0.769	0.957	0.741	0.954	1.025	0.610	0.808	0.548
60–64	1.087	1.297	1.072	0.852	0.933	1.218	0.701	0.955	0.551
65–69	1.093	1.150	1.226	0.954	1.067	1.366	0.901	1.089	0.508
70–74	0.988	1.507	1.375	1.076	1.092	1.740	1.219	1.369	0.508
75–79	1.425	2.241	1.523	1.219	1.215	2.259	1.629	1.650	1.083
80–84	1.620	1.909	1.714	1.372	1.513	2.729	2.570	1.650	1.876
85+	2.008	2.738	1.914	1.555	2.438	3.358	3.705	1.650	1.876

The specification of relative interval survival ($RIS_{\alpha,t,\tau}$) for age α, calendar year t across the interval (τ - 1) since diagnosis in years, separately for each cancer site, was of the form:

$$RIS_{\alpha,t,\tau} = 1 - (1 - RIS_1)\, A_\alpha\, T_t\, Y_\tau \quad (1)$$

where RIS_1 is the relative interval survival after 1 year for all ages, averaged across the calendar years 1973 to 1997, A_α is the ratio of the relative probability of death after 1 year at age α to the relative probability of death after 1 year for all ages averaged across the calendar years 1973 to 1997 $\left(A_\alpha = \dfrac{1 - RIS_{\alpha,1973-97}}{1 - RIS_1} \right)$, T_t is the ratio of the relative probability of death after 1 year for all ages in calendar year t to the relative probability of death after 1 year for all ages from 1973 to 1997 $\left(T_t = \dfrac{1 - RIS_t}{1 - RIS_1} \right)$ and Y_τ is the ratio of the relative probability of death after t years for all ages to the relative probability of death after 1 year for all ages from 1973 to 1997 $\left(Y_\tau = \dfrac{1 - RIS_{\tau,1973-97}}{1 - RIS_1} \right)$.

To estimate survival for developing regions where little or no data is available, we replaced T_t for the "equivalent" calendar year survival term T_r in the equation (1) for each region. T_r is the ratio of the relative probability of death after 1 year for all ages in the relevant region to the relative probability of death after 1 year for all ages in the SEER data, averaged across the calendar years 1973 to 1997. In this way, we obtain a new calendar year survival term for the model.

Equivalent period survival terms were estimated for each region by examining the relationship between period survival terms and gross domestic product per capita (measured in international dollars adjusted for purchasing power parity) using the following data: 1) SEER survival data for the USA for the years 1973 to 1997 [11,29]; 2)

Connecticut survival data for the years 1950 and 1958 [30]; 3) survival data for the late 1980s from cancer registries in 5 developing countries [31]; 4) survival data for European countries [32]; 5) specific recent national estimates of cancer survival as published [24,33]. Survivorship functions were estimated to derive regional relative survival from registry data by fitting a Weibull distribution function. To allow for a proportion who are cured and never die from the cancer, we modify the Weibull model as follows:

$$S(t) = k + (1 - k)\, \exp\left(-(\lambda\, t)^\gamma \right) \quad (2)$$

where k is the proportion who never die from the cancer, λ is the location parameter ($1/\lambda$ is the time at which 50% of those will die have died) and γ is the shape parameter. The mean survival time for those who die ($\overline{T}$) is given by

$$\overline{T} = \frac{1}{\lambda} \Gamma\left(\frac{\gamma + 1}{\gamma} \right) \quad (3)$$

where Γ denotes the gamma distribution. The analysis of survival data in developed regions suggested that the 10-year relative survival can be used as an estimate of the proportion who never die from the cancer. This is particularly useful when relative survival point estimates fluctuate significantly and the plausible exact solutions did not obtained as in some developing regions. To assess the goodness-of-fit of the survival curve, we compared fitted 5-year survival with the observed survival and ensured the good fits in all cases.

There are substantial variations in relative interval survival (all ages) among countries; these variations are even larger, and fluctuate substantially with age, when the age-sex specific survival estimates are examined. There is also a possibility of correlation among the observations within a same region. Thus we employed both liner and non-linear feasible generalised least squares (FGLS) by maximum likelihood estimation to accommodate with hetero-

Table 4: Estimated time since diagnosis (cohort) parameters (Y_τ) by cancer site, year from diagnosis and sex

Year	Oral Male	Oral Female	Oesophagus Male	Oesophagus Female	Stomach Male	Stomach Female	Colon and rectum Male	Colon and rectum Female	Liver Male	Liver Female	Pancreas Male	Pancreas Female
1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	0.877	0.855	0.752	0.753	0.640	0.564	0.609	0.572	0.591	0.542	0.721	0.720
3	0.542	0.472	0.462	0.483	0.387	0.349	0.443	0.388	0.368	0.352	0.434	0.430
4	0.377	0.329	0.313	0.251	0.251	0.186	0.345	0.267	0.265	0.219	0.257	0.226
5	0.300	0.268	0.214	0.216	0.173	0.117	0.240	0.189	0.268	0.174	0.156	0.170
6	0.235	0.282	0.175	0.181	0.109	0.099	0.185	0.134	0.116	0.104	0.150	0.108
7	0.241	0.268	0.173	0.109	0.087	0.065	0.144	0.114	0.118	0.083	0.116	0.092
8	0.217	0.224	0.145	0.168	0.063	0.084	0.094	0.085	0.073	0.093	0.100	0.104
9	0.211	0.241	0.095	0.168	0.035	0.068	0.079	0.068	0.041	0.075	0.062	0.078
10	0.236	0.213	0.084	0.138	0.048	0.046	0.068	0.059	0.000	0.049	0.085	0.051
11	0.219	0.197	0.057	0.145	0.016	0.041	0.024	0.033	0.000	0.045	0.047	0.048
12	0.217	0.180	0.030	0.142	0.000	0.031	0.000	0.014	0.000	0.034	0.029	0.036
13	0.214	0.164	0.004	0.139	0.000	0.021	0.000	0.000	0.000	0.022	0.010	0.023
14	0.211	0.147	0.000	0.137	0.000	0.010	0.000	0.000	0.000	0.010	0.000	0.010
15	0.208	0.131	0.000	0.134	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Year	Lung Male	Lung Female	Bladder Male	Bladder Female	Non-Hodgkin lymphoma Male	Non-Hodgkin lymphoma Female	Hodgkin lymphoma Male	Hodgkin lymphoma Female	Multiple myeloma Male	Multiple myeloma Female	Leukaemia Male	Leukaemia Female
1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	0.683	0.693	0.621	0.478	0.442	0.480	0.589	0.654	0.725	0.755	0.466	0.438
3	0.402	0.384	0.391	0.267	0.277	0.289	0.411	0.417	0.792	0.711	0.334	0.280
4	0.254	0.235	0.313	0.191	0.235	0.241	0.384	0.290	0.728	0.692	0.282	0.259
5	0.181	0.186	0.277	0.141	0.220	0.233	0.304	0.224	0.764	0.732	0.251	0.236
6	0.145	0.144	0.251	0.141	0.195	0.233	0.301	0.290	0.689	0.740	0.225	0.206
7	0.123	0.134	0.249	0.131	0.188	0.222	0.307	0.263	0.643	0.727	0.225	0.189
8	0.120	0.117	0.231	0.108	0.166	0.217	0.252	0.219	0.601	0.579	0.215	0.172
9	0.113	0.116	0.252	0.139	0.163	0.183	0.242	0.191	0.596	0.570	0.194	0.165
10	0.113	0.114	0.257	0.165	0.145	0.171	0.198	0.188	0.510	0.453	0.168	0.153
11	0.101	0.102	0.253	0.154	0.137	0.177	0.179	0.124	0.537	0.466	0.162	0.139
12	0.093	0.094	0.255	0.160	0.128	0.170	0.161	0.102	0.519	0.430	0.147	0.126
13	0.085	0.087	0.256	0.165	0.119	0.164	0.144	0.082	0.501	0.397	0.133	0.113
14	0.078	0.079	0.258	0.171	0.112	0.157	0.128	0.064	0.485	0.367	0.118	0.100
15	0.070	0.071	0.260	0.177	0.104	0.152	0.114	0.046	0.471	0.338	0.104	0.088

Year	Melanoma Male	Melanoma Female	Miscellaneous Male	Miscellaneous Female	Breast Female	Cervix Female	Uterus Female	Ovary Female	Prostate Male
1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	0.967	0.972	0.423	0.361	1.205	0.869	0.656	0.726	1.108
3	0.801	0.962	0.215	0.194	1.297	0.507	0.418	0.461	1.142
4	0.622	0.741	0.152	0.127	1.154	0.337	0.233	0.315	1.065
5	0.459	0.514	0.109	0.091	1.051	0.232	0.141	0.232	1.051
6	0.392	0.531	0.097	0.079	0.926	0.196	0.118	0.169	1.038

Table 4: Estimated time since diagnosis (cohort) parameters (Y_τ) by cancer site, year from diagnosis and sex *(Continued)*

7	0.280	0.339	0.082	0.067	0.879	0.162	0.070	0.110	1.113
8	0.239	0.329	0.070	0.056	0.744	0.146	0.066	0.113	1.099
9	0.220	0.248	0.069	0.052	0.756	0.095	0.048	0.089	1.022
10	0.085	0.217	0.066	0.050	0.705	0.119	0.038	0.061	1.013
11	0.101	0.117	0.054	0.039	0.633	0.078	0.014	0.038	1.015
12	0.026	0.045	0.047	0.032	0.576	0.056	0.000	0.014	1.001
13	0.000	0.000	0.039	0.025	0.520	0.034	0.000	0.000	0.987
14	0.000	0.000	0.032	0.017	0.464	0.012	0.000	0.000	0.973
15	0.000	0.000	0.024	0.010	0.407	0.000	0.000	0.000	0.959

scedasticity and correlation among the observations, and chose the best fit model for each site [34]. We did not include region fixed-effects since the data are not available for all sub regions.

Model estimation of cancer death distribution

For the estimation of the number of deaths and cancer death distribution by site, we needed to estimate the number of individuals who survived up to 2000 by age and time of survival as well as their corresponding probability of death during this year. The number of surviving individuals at age α in 2000 was calculated by multiplying incidence at age α in year $(2000 - \tau)$ by the observed interval survival for τ years since diagnosis for individuals aged α in 2000 ($OIS_{\alpha,\tau}$), and summing over τ. To estimate OIS$_{\alpha,\tau}$, we first calculated the relative cumulative survival ($RCS_{\alpha,\tau}$) for every single age and year of survival for 2000, by multiplying $RIS_{\alpha,\tau}$ over the years of survival. In a standard life table format, $OIS_{\alpha,\tau}$ is written of the from:

$$OIS_{\alpha,\tau} = RCS_{\alpha,\tau}\left(l_{\alpha+1}/l_{\alpha-\tau+1}\right) = \exp\left(-\left(-\ln\left(RCS_{\alpha,\tau}\right) + \sum_{x=\alpha-\tau+1}^{\alpha} h_x\right)\right) \quad (4)$$

where l_x is the number of individuals surviving at exact age x in the life table, $h_x = \ln(l_{x+1}/l_x)$, α is age and τ *is* time since diagnosis.

The number of individuals $S_{\alpha,\tau}$ who had survived up to 2000 was by multiplying incidence and observed interval survival for the corresponding year of age and survival time:

$$S_{\alpha,\tau} = I_{\alpha-\tau,2000-\tau}OIS_{\alpha,\tau} \quad (5)$$

where $I_{\alpha,t}$ is the incidence at age α in calendar year t. For example, the number of individuals who were 7 years of age ($\alpha = 7$) in 2000, and who had survived cancer for 4 years ($\tau = 4$) in 2000 was calculated by multiplying the incidence of cancer for the cohort of individuals who were 3 years of age ($\alpha - \tau = 3$) in 1996 (= 2000 - τ) (year of diag-

nosis) by the $OIS_{\alpha,\tau}$ calculated for a 7 year old person who had survived 4 years since cancer diagnosis.

The probability of dying in 2000 due to cancer hazard, for each single age and year of survival, was calculated as follows:

$$P_{\alpha,\tau} = [1 - \exp(-(-\ln(RIS_{\alpha,\tau}) + h_\alpha))]\ [-\ln(RIS_{\alpha,\tau})/(-\ln(RIS_{\alpha,\tau}) + h_\alpha)] \quad (6)$$

For each site, the number of deaths in 2000, among those individuals aged α years and who had survived cancer for τ years, was estimated by multiplied the number of survivors $S_{\alpha,\tau}$ by the relevant probability of dying in 2000 due to cancer hazard $P_{\alpha,\tau}$. The total cancer deaths of the same site in 2000 at age α years is then estimated by summing over all survival times τ:

$$D_\alpha = \sum_\tau S_{\alpha,\tau}P_{\alpha,\tau} \quad (7)$$

Based on these region-specific adjusted incidence estimates and survival levels, cancer deaths were calculated by equations (3)-(6) for each region by age group and sex to estimate the distribution, but not the magnitude, of cancer by site, sex, and age group.

Validation of the model

We tested the validity and performance of the proposed survival model in three different ways. Firstly, we compared our estimated $RIS_{\alpha,t,\tau}$ for $\tau = 1$ to 10 years individuals diagnosed with cancer in 1986 with the SEER $RIS_{\alpha,t,\tau}$ for $\tau = 1$ to 10 years for the same cohort of individuals. Secondly, we compared the model estimates of cancer mortality distribution with the observed distributions in the regions with good vital records (AmrB, EurA, EurB, EurC and WprA sub regions). AmrA sub region was excluded for the validation purpose since it includes the United States. Finally we compared the cancer death distribution of our model with the Globocan 2000 estimates for the regions where no vital records are available (AfrD,

Table 5: Estimated period parameters (T_t) for 1981–1995 by cancer site, calendar year, and sex

Year	Oral Male	Oral Female	Oesophagus Male	Oesophagus Female	Stomach Male	Stomach Female	Colon and rectum Male	Colon and rectum Female	Liver Male	Liver Female	Pancreas Male	Pancreas Female
1981	0.991	1.035	1.121	1.180	1.060	1.003	0.983	1.013	1.092	1.092	1.004	1.022
1982	1.019	1.048	1.054	0.971	0.949	1.066	0.986	1.017	1.031	1.036	0.998	1.014
1983	1.065	1.095	1.046	1.046	0.939	0.992	0.989	1.026	1.019	1.033	0.997	0.980
1984	0.925	1.090	0.959	0.944	1.051	0.967	0.998	0.938	0.991	1.053	1.039	0.989
1985	0.895	1.053	0.975	0.957	0.993	0.994	0.904	0.949	0.974	1.000	1.018	1.022
1986	1.070	0.910	0.997	0.995	1.007	0.900	0.850	0.893	1.015	1.028	0.988	1.000
1987	0.994	0.979	0.996	0.954	1.002	0.976	0.883	0.904	0.952	0.966	0.988	0.987
1988	0.954	0.887	0.998	0.925	0.940	0.918	0.914	0.917	1.009	1.030	0.976	1.008
1989	1.104	0.984	0.933	0.943	0.919	0.980	0.870	0.876	0.944	1.029	0.984	0.967
1990	1.000	0.896	0.886	1.049	0.937	0.956	0.824	0.867	0.959	0.985	0.972	0.986
1991	0.991	0.959	0.880	0.893	0.998	0.932	0.810	0.824	0.969	0.944	0.962	0.974
1992	1.070	0.992	0.889	0.902	0.952	0.903	0.832	0.871	1.011	0.959	0.993	0.949
1993	0.998	0.865	0.904	1.017	0.966	0.944	0.888	0.894	0.972	0.982	0.994	0.985
1994	0.842	0.965	0.936	0.938	0.935	0.966	0.833	0.875	0.933	0.914	0.964	0.983
1995	0.899	0.942	0.887	0.927	0.900	0.981	0.896	0.971	0.942	0.943	0.995	0.987

Year	Lung Male	Lung Female	Bladder Male	Bladder Female	Non-Hodgkin lymphoma Male	Non-Hodgkin lymphoma Female	Hodgkin lymphoma Male	Hodgkin lymphoma Female	Multiple myeloma Male	Multiple myeloma Female	Leukaemia Male	Leukaemia Female
1981	1.020	1.017	0.916	1.052	0.882	0.960	0.994	1.675	0.985	1.107	1.004	0.916
1982	0.994	1.000	1.179	0.775	1.033	1.069	1.240	0.888	1.008	0.837	1.042	1.043
1983	0.988	0.996	0.998	1.005	0.788	0.946	0.840	1.009	0.809	0.975	1.033	1.030
1984	0.994	1.009	1.090	0.762	0.883	0.999	0.756	0.775	1.066	1.018	0.981	0.985
1985	0.997	0.972	0.972	1.039	0.949	1.021	0.834	0.947	0.914	0.989	0.962	0.963
1986	0.976	0.986	0.925	0.883	1.010	1.102	0.994	0.579	1.015	0.911	0.982	0.936
1987	0.977	1.015	0.885	0.995	1.016	0.898	0.867	0.880	0.795	0.876	0.959	1.023
1988	0.980	0.995	0.851	0.859	1.043	0.967	0.854	0.978	0.866	0.930	0.905	0.937
1989	0.982	0.981	0.746	0.908	1.064	0.997	1.036	0.747	0.969	0.966	0.977	0.946
1990	0.982	0.985	0.869	1.040	1.125	0.959	0.863	1.021	0.996	0.938	0.918	0.930
1991	0.968	0.971	0.918	0.907	1.115	0.991	0.634	0.811	0.782	0.880	0.930	0.959
1992	0.964	0.986	0.894	0.942	1.059	1.000	0.831	0.579	0.946	1.176	0.920	0.925
1993	0.979	0.977	0.844	0.897	1.069	0.965	0.998	0.583	1.036	0.834	0.927	0.972
1994	0.975	0.990	0.844	0.935	1.091	0.987	0.590	0.787	0.892	0.965	0.927	1.018
1995	0.968	0.990	0.815	0.878	1.116	1.007	0.719	0.655	0.895	0.936	0.952	1.026

Year	Melanoma Male	Melanoma Female	Miscellaneous Male	Miscellaneous Female	Breast Female	Cervix Female	Uterus Female	Ovary Female	Prostate Male
1981	1.230	1.467	0.928	1.089	1.182	1.091	1.229	1.076	1.589
1982	1.332	0.765	0.921	1.060	1.131	1.020	1.160	1.036	1.556
1983	0.830	0.689	0.894	1.044	1.174	0.862	1.104	1.039	1.454
1984	1.124	0.567	0.902	0.996	1.095	0.921	1.160	0.931	1.411
1985	1.014	0.979	0.855	0.994	1.056	1.041	0.932	0.972	1.301
1986	0.838	0.830	0.862	0.966	0.949	1.132	0.953	1.084	1.358

Table 5: Estimated period parameters (T_t) for 1981–1995 by cancer site, calendar year, and sex *(Continued)*

1987	0.971	1.038	0.843	0.982	0.810	0.993	1.089	1.015	1.210
1988	0.784	0.633	0.871	0.977	0.810	1.002	1.056	0.926	0.976
1989	1.124	0.969	0.823	0.954	0.764	0.870	1.047	0.849	0.833
1990	0.925	0.945	0.784	0.941	0.779	0.941	1.001	0.833	0.651
1991	0.757	0.574	0.770	0.894	0.695	0.881	1.075	0.817	0.309
1992	0.838	1.197	0.806	0.897	0.667	1.002	0.955	0.738	0.278
1993	0.902	0.813	0.803	0.910	0.723	0.848	1.042	0.827	0.226
1994	0.610	1.163	0.777	0.891	0.649	0.985	1.115	0.836	0.299
1995	0.757	0.543	0.805	0.910	0.721	0.862	0.968	0.779	0.290

AfrE, SearB and SearD sub regions) to assess whether our model estimates are comparable to the estimates extrapolated from the actual observed data from the registries. In all cases, non-parametric tests for trends and Pearson's correlation were performed to examine whether the models estimates and observed data are consistent with each other. All statistical analysis was performed by STATA 7.0 (STATA corporation, College Station, TX).

Results

Parameter estimates

Relative interval survival (RIS) was directly obtained from the SEER database for every age group, sex, and cancer site The probability of death in the first year (1 - RIS_1), which is most crucial for the survival of most cancer patients, is shown in Table 2. The probability of death varied significantly from less than 5% in cancers of melanoma, breast, uterus, and prostate to over 80% in pancreas and liver cancers.

Relative interval survival (*RIS*) was further indexed to the three parameters in the multiplicative cancer survival model by dividing each of the time-specific probability of death (1 - *RIS*) by the probability of death at 1-year interval (1 - RIS_1). Tables 3,4,5 represent the indexed estimates of three parameters by site for every 5-year age group from 0 up to 85+ years of age, and for calendar year from 1981 to 1995, and for time since cancer diagnosis from 1- up to 15-year survival, respectively. While there is considerable variation in the cohort parameters which reflect the prognosis among patients since the time of diagnosis, both age and period parameters are generally consistent across different types of cancer.

Based on the fitted data for each site and sex, and the estimated GDP per capita in international dollars for each region in 2000, T_r factors were estimated for each site and sex for each GBD 2000 region. The results are presented in Table 6. An example is shown for breast cancer: knowing that GDP per capita in AfrD was $1,158 in 2000, this corresponded to an indexed calendar year-specific T_t = 2.748. This was then the value used in the age-period-cohort sur-

vival model for breast cancer in the AfrD sub region. A similar process was applied to the other regions and for other cancer sites.

The period parameters (T_r) for all the available survival data and fitted values from regression analysis were plotted against GDP per capita (international dollars) for each site and sex as shown in Figures 1,2,3,4,5,6. The largest variation in survival was observed among cancers of breast, melanoma of the skin, and hematological malignancies such as lymphoma and leukemia. For the cancers of cervix and colon and rectum, both early detection and availability of treatment determine the survival and the variation among regions were moderate. On the other hand, cancers with very poor prognosis such as liver, lung, and pancreas cancers showed very small variations across the regions regardless of the levels of national income.

Model performance and validation

In order to check the performance of the model, we graphically compared our estimated $RIS_{\alpha,t,\tau}$ for τ = 1 to 10 years individuals diagnosed with cancer in 1986 with the SEER $RIS_{\alpha,t,\tau}$ for τ = 1 to 10 years for the same cohort of individuals. We show the results obtained for males and females 55–59 years old, and for every cancer site in Figures 7,8,9. From these figures, we can observe that the model predicts very well the relative interval survivals. For those cancer sites with greater number of cases, such as colon, lung, breast, corpus uteri, and prostate cancer, the model fits very well. For those with smaller numbers such as cancers of liver and pancreas, the estimated *RIS* smoothes the curves for the observed *RIS*, also showing a very good fit.

We also tested the validity of our model when applied to other populations. We have chosen the age groups from 45 to 79 in which cancer mortality rate is relatively stable and the probability of miscoding of cause of death is small. Figures 10,11 shows the comparison between model estimates and vital registration data for six sub regions in age group 65–69 (AmrB, EurA, EurB1, EurB2, EurC, and WprA sub regions). The estimated coefficients and p-values for the test of Pearson's correlation by using all data
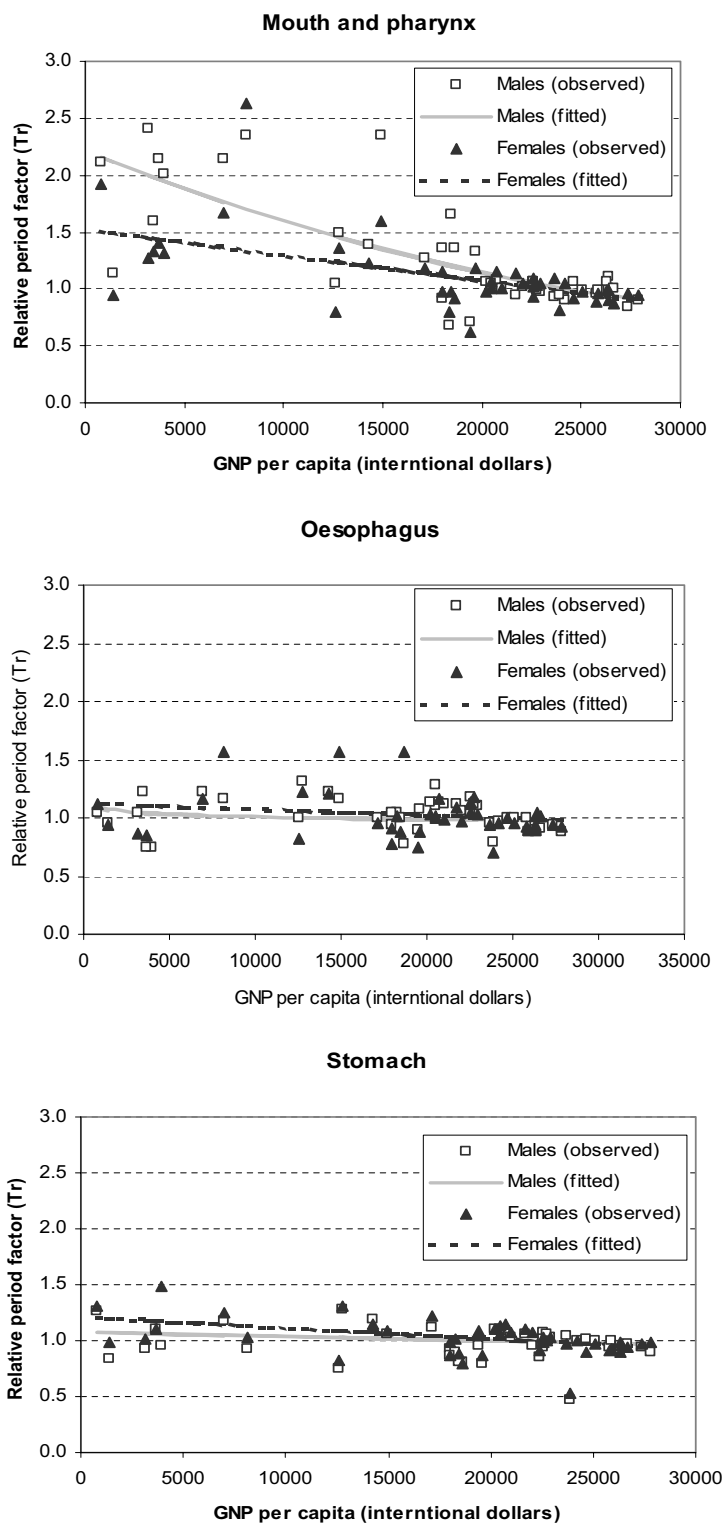
Figure 1
Observed and fitted period factor by region (T_r) versus GDP per capita (1)

Table 6: Estimated regional period parameters (T_r) for 2000 by cancer site and sex

Region	Oral		Oesophagus		Stomach		Colon and rectum		Liver		Pancreas	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
AfrD	2.132	1.493	1.072	1.119	1.070	1.198	1.892	1.794	1.050	1.102	1.003	1.049
AfrE	2.111	1.486	1.065	1.117	1.069	1.195	1.880	1.784	1.050	1.101	1.003	1.048
AmrA	0.887	0.873	0.975	0.973	0.965	0.930	0.845	0.859	0.985	1.000	1.002	0.994
AmrB	1.747	1.357	1.017	1.087	1.047	1.140	1.663	1.590	1.046	1.080	1.003	1.037
AmrD	1.984	1.443	1.040	1.107	1.061	1.177	1.809	1.720	1.049	1.094	1.003	1.044
EmrB	1.774	1.368	1.019	1.089	1.049	1.144	1.681	1.606	1.046	1.082	1.003	1.038
EmrD	2.005	1.451	1.043	1.109	1.063	1.180	1.821	1.731	1.049	1.095	1.003	1.045
EurA	1.122	1.059	0.985	1.017	0.997	1.011	1.160	1.141	1.017	1.030	1.002	1.010
EurB1	1.797	1.376	1.021	1.091	1.050	1.148	1.695	1.618	1.047	1.083	1.003	1.038
EurB2	2.069	1.472	1.054	1.114	1.066	1.189	1.857	1.763	1.050	1.099	1.003	1.047
EurC	1.868	1.402	1.027	1.097	1.055	1.159	1.739	1.658	1.048	1.087	1.003	1.041
SearB	1.933	1.426	1.033	1.103	1.058	1.169	1.778	1.693	1.049	1.091	1.003	1.043
SearD	2.127	1.491	1.071	1.118	1.070	1.197	1.890	1.792	1.050	1.102	1.003	1.048
WprA	0.979	0.958	0.979	0.993	0.980	0.967	0.989	0.987	1.001	1.014	1.002	1.001
WprB1	1.990	1.445	1.041	1.108	1.062	1.178	1.812	1.723	1.049	1.095	1.003	1.044
WprB2	2.095	1.481	1.061	1.116	1.068	1.193	1.872	1.776	1.050	1.100	1.003	1.048
WprB3	1.967	1.437	1.038	1.106	1.060	1.174	1.799	1.711	1.049	1.093	1.003	1.044

Region	Lung		Bladder		Non-Hodgkin lymphoma		Hodgkin lymphoma		Multiple myeloma		Leukaemia	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
AfrD	1.141	1.212	3.526	2.844	1.466	1.581	2.176	2.419	1.081	1.213	1.539	1.493
AfrE	1.139	1.210	3.492	2.819	1.459	1.574	2.160	2.399	1.080	1.209	1.532	1.486
AmrA	0.984	0.995	0.493	0.653	0.891	0.913	0.676	0.557	0.945	0.907	0.880	0.875
AmrB	1.107	1.165	2.864	2.365	1.340	1.435	1.849	2.013	1.052	1.146	1.395	1.358
AmrD	1.129	1.195	3.284	2.669	1.420	1.528	2.057	2.271	1.070	1.188	1.487	1.444
EmrB	1.110	1.169	2.915	2.402	1.350	1.447	1.874	2.044	1.054	1.151	1.406	1.368
EmrD	1.131	1.198	3.320	2.695	1.427	1.536	2.075	2.293	1.072	1.192	1.494	1.451
EurA	1.031	1.061	1.405	1.312	1.064	1.114	1.127	1.117	0.986	0.999	1.078	1.061
EurB1	1.112	1.172	2.956	2.432	1.358	1.456	1.894	2.069	1.056	1.155	1.415	1.377
EurB2	1.136	1.205	3.424	2.771	1.447	1.559	2.126	2.357	1.077	1.203	1.517	1.472
EurC	1.118	1.181	3.083	2.524	1.382	1.484	1.957	2.148	1.061	1.168	1.443	1.403
SearB	1.124	1.189	3.197	2.606	1.403	1.509	2.014	2.217	1.067	1.180	1.468	1.426
SearD	1.141	1.212	3.519	2.839	1.464	1.580	2.173	2.415	1.081	1.212	1.538	1.492
WprA	1.006	1.025	0.908	0.953	0.970	1.004	0.881	0.812	0.964	0.949	0.970	0.959
WprB1	1.129	1.196	3.294	2.676	1.422	1.530	2.062	2.277	1.071	1.189	1.489	1.446
WprB2	1.138	1.208	3.468	2.802	1.455	1.569	2.148	2.384	1.079	1.207	1.527	1.481
WprB3	1.127	1.193	3.255	2.648	1.414	1.522	2.043	2.253	1.069	1.186	1.480	1.438

Region	Melanoma		Miscellaneous		Breast	Cervix	Uterus	Ovary	Prostate	GDP Per capita
	Male	Female	Male	Female	Female	Female	Female	Female	Male	
AfrD	3.811	6.596	1.720	1.741	2.748	1.647	1.896	1.261	5.559	1,158
AfrE	3.775	6.174	1.709	1.731	2.725	1.638	1.885	1.247	5.505	1,465
AmrA	0.452	0.819	0.718	0.849	0.657	0.845	0.881	1.064	0.647	28,910
AmrB	3.093	3.311	1.501	1.546	2.292	1.472	1.674	1.149	4.487	7,217
AmrD	3.550	4.680	1.640	1.670	2.582	1.583	1.815	1.196	5.168	3,367
EmrB	3.149	3.431	1.518	1.561	2.327	1.485	1.692	1.153	4.570	6,749
EmrD	3.589	4.864	1.652	1.680	2.606	1.592	1.827	1.202	5.226	3,040
EurA	1.481	1.431	1.019	1.117	1.286	1.086	1.186	1.085	2.125	20,562

Table 6: Estimated regional period parameters (T_r) for 2000 by cancer site and sex *(Continued)*

EurB1	3.194	3.535	1.532	1.573	2.355	1.496	1.705	1.157	4.637	6,371
EurB2	3.702	5.541	1.687	1.711	2.678	1.620	1.862	1.225	5.395	2,085
EurC	3.332	3.897	1.574	1.611	2.443	1.530	1.748	1.169	4.843	5,206
SearB	3.455	4.297	1.611	1.644	2.521	1.560	1.786	1.183	5.027	4,168
SearD	3.804	6.506	1.718	1.739	2.744	1.645	1.894	1.258	5.549	1,218
WprA	0.922	1.072	0.855	0.971	0.943	0.954	1.020	1.073	1.319	25,115
WprB1	3.561	4.728	1.644	1.673	2.589	1.585	1.819	1.198	5.184	3,277
WprB2	3.749	5.920	1.701	1.724	2.708	1.631	1.877	1.238	5.466	1,688
WprB3	3.519	4.545	1.631	1.661	2.562	1.575	1.806	1.191	5.122	3,631

points in age groups from 45 to 79 were also presented. When compared with site-specific mortality distribution of the Globocan 2000 based on regional cancer registries for AFRO and SEARO regions, the model estimate also yielded consistent mortality distribution pattern (Figure 12). In all cases, the correlation coefficients were in the range of 0.91 to 0.98, suggesting that model estimates for these regions are quite consistent with the observed cancer mortality distribution.

Probability of 5-year survival and mean duration by site
The proposed model also yields RIS for various years and the mean duration of cancer by site, both of which are important inputs for the future estimation of cancer morbidity burden in terms of years lived with disability (YLDs). As an illustration of the further use of our model, Figures 13 and 14 show, respectively, the conventional 5-year survival and average duration for the female cancer patients aged 45–54 in four different sub regions (AfrE, AmrA, SearD and WprB3). Depending on the site, the chance of 5-year survival and average duration varied considerably across the resigns, which are consistent with the estimated survival pattern above.

Discussion
As a part of the Global Burden of Disease 2000 (GBD 2000) study, we have developed a multiplicative model of relative interval survival for cancer by site based on the best available evidence from published population-based survival data from both developed and developing countries. Because of the sparseness of survival data for the developing regions of the world, we decided to use all the available data, including the most valuable cancer registry data compiled by the International Agency for Research on Cancer (IARC), national cancer registries, and lengthy time series data from the United States, to establish trends in survival with gross domestic product (GDP) per capita and then to use latest estimates of GDP per capita for developing regions, in order to estimate survival by site.

This approach takes into account, through increases in average GDP per capita for regions, the likely improvements

in survival over the periods since those for which developing country survival data are available. Since our survival model adjusted for age and differences in competing mortality in each population by employing relative interval survival, the remaining variations in survival are likely to be due to differences in diagnosis and availability of appropriate treatment options. For instance, the large variation in survival was observed among cancers of bladder, breast, melanoma of the skin, and hematological malignancies such as lymphoma and leukemia for which effective therapy is established in developed regions. For the cancers of cervix and colon and rectum, both early detection and availability of treatment determine the survival and the variation among regions were moderate. On the other hand, cancers with very poor prognosis showed very small variations across the regions. The survival pattern across the regions are consistent with previous analysis based on the IARC cancer registry data [35].

The proposed model takes into account time in its three dimensions: age, calendar year (period) and time since cancer (cohort) and, due to the availability of data, the model outcome was compared the to the data reported by the US vital statistics and other regions of the world. This has given us the opportunity to evaluate our model and the data available.
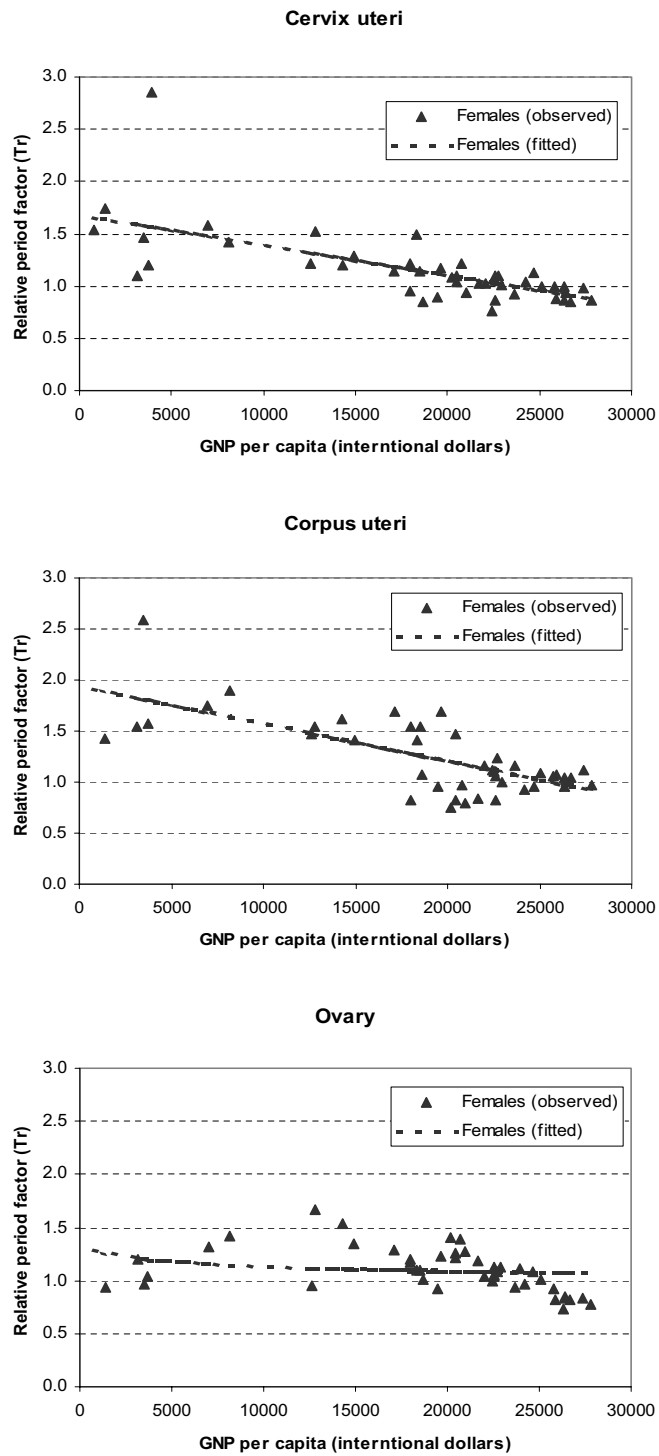
However, perhaps the main advantage of this approach to estimating regional survival distributions by cancer site for developing regions is that the model correctly estimates survival and smoothes it and ensures that regional survival estimates are consistent with trends in survival across all regions, where the numbers for some cancer sites are small and highly fluctuate, and are inconsistent with other regions. For example, as can be seen in Figure 1a, cancer registry survival estimates for some sites in some developing countries are better than recent experience in the United States, or significantly below the trend line with GDP per capita, suggesting that survival may have been overestimated due to small numbers or incomplete case follow-up. In these cases, the survival model provides survival estimates more consistent with the com-

Colon and rectum cancer

Liver cancer

Pancreas
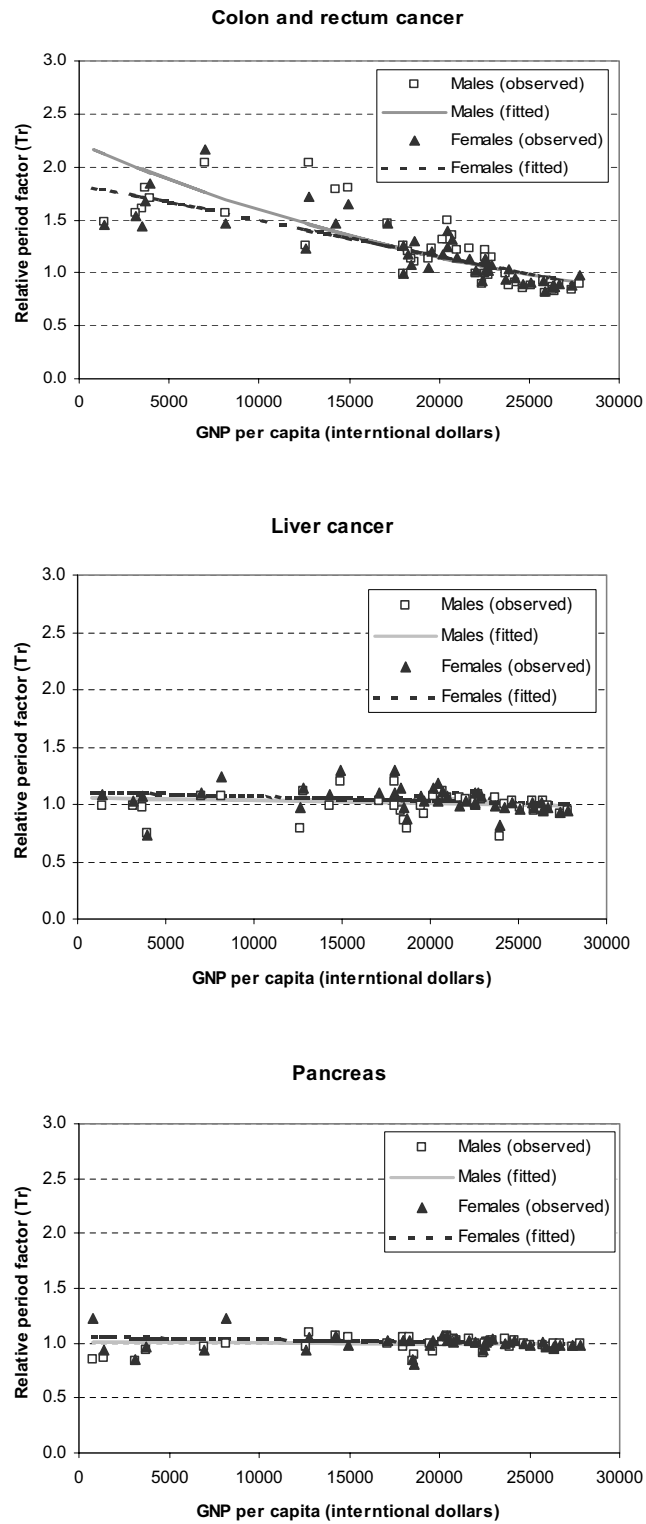
Figure 2
Observed and fitted period factor by region (T_r) versus GDP per capita (2)

Trachea, bronchus and lung

Melanoma of the skin

Breast
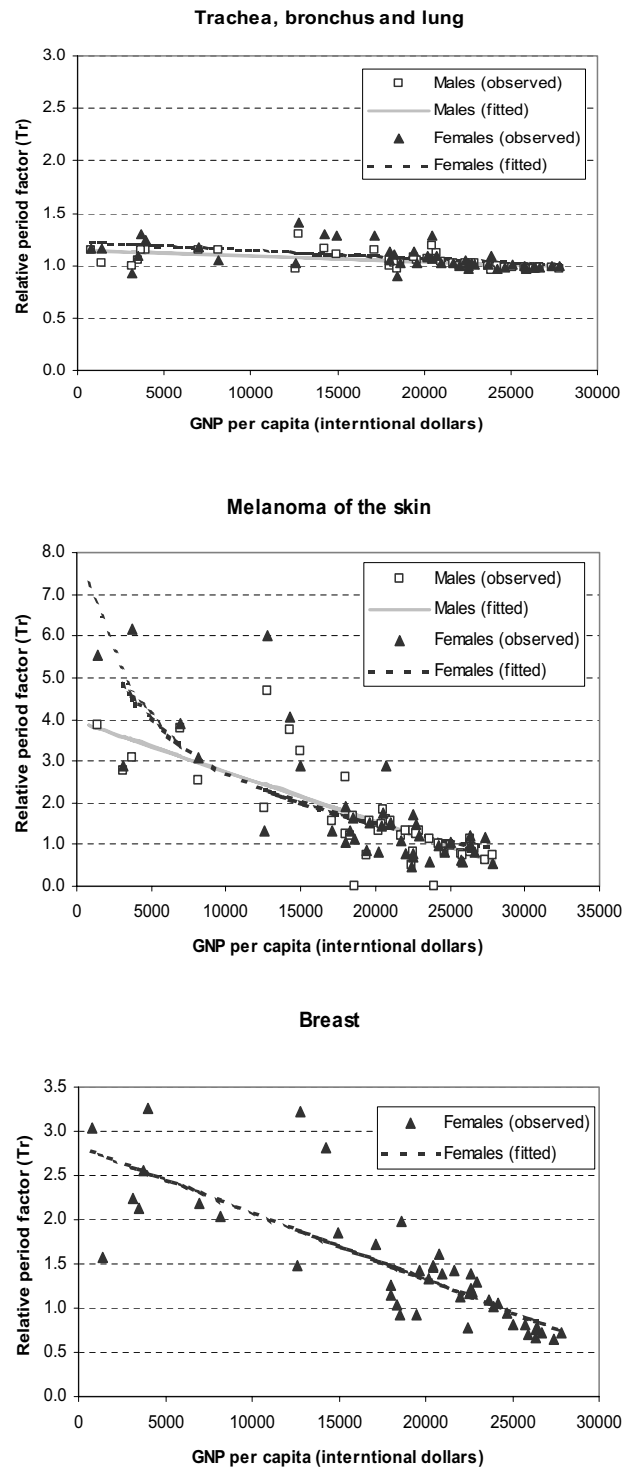
Figure 3
Observed and fitted period factor by region (T_r) versus GDP per capita (3)

Cervix uteri

Corpus uteri

Ovary

Figure 4
Observed and fitted period factor by region (T_r) versus GDP per capita (4)

Prostate

Non-Hodgkin lymphoma

Bladder
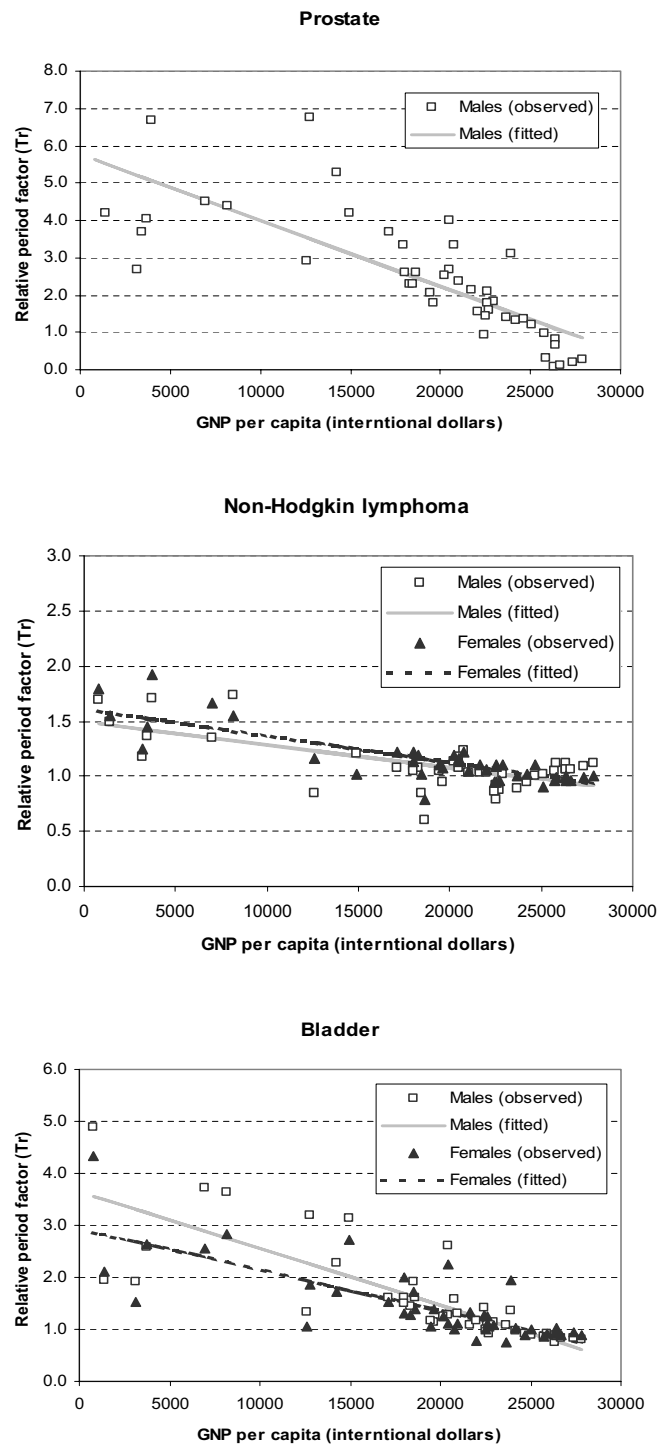
Figure 5
Observed and fitted period factor by region (*T$_r$*) versus GDP per capita (5)

Hodgkin lymphoma

Multiple myeloma

Leukaemia
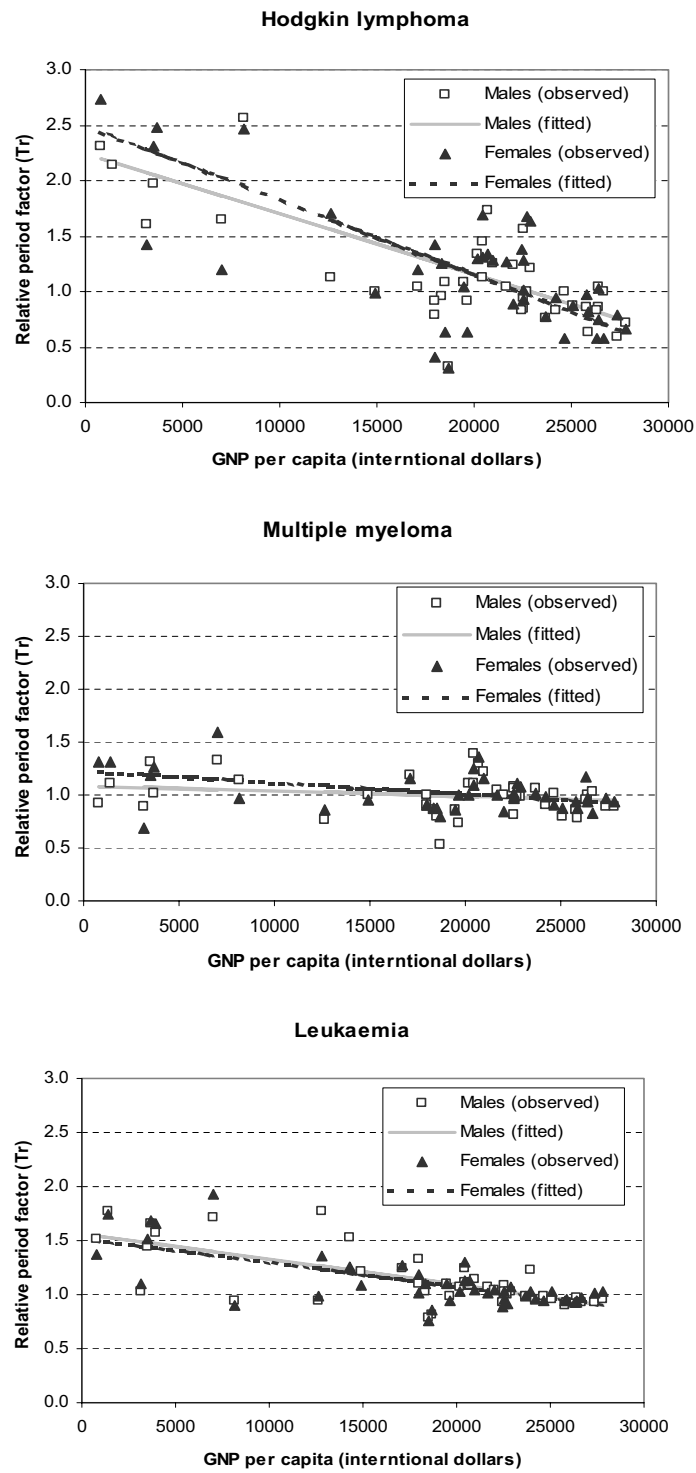
Figure 6
Observed and fitted period factor by region (*T$_r$*) versus GDP per capita (6)
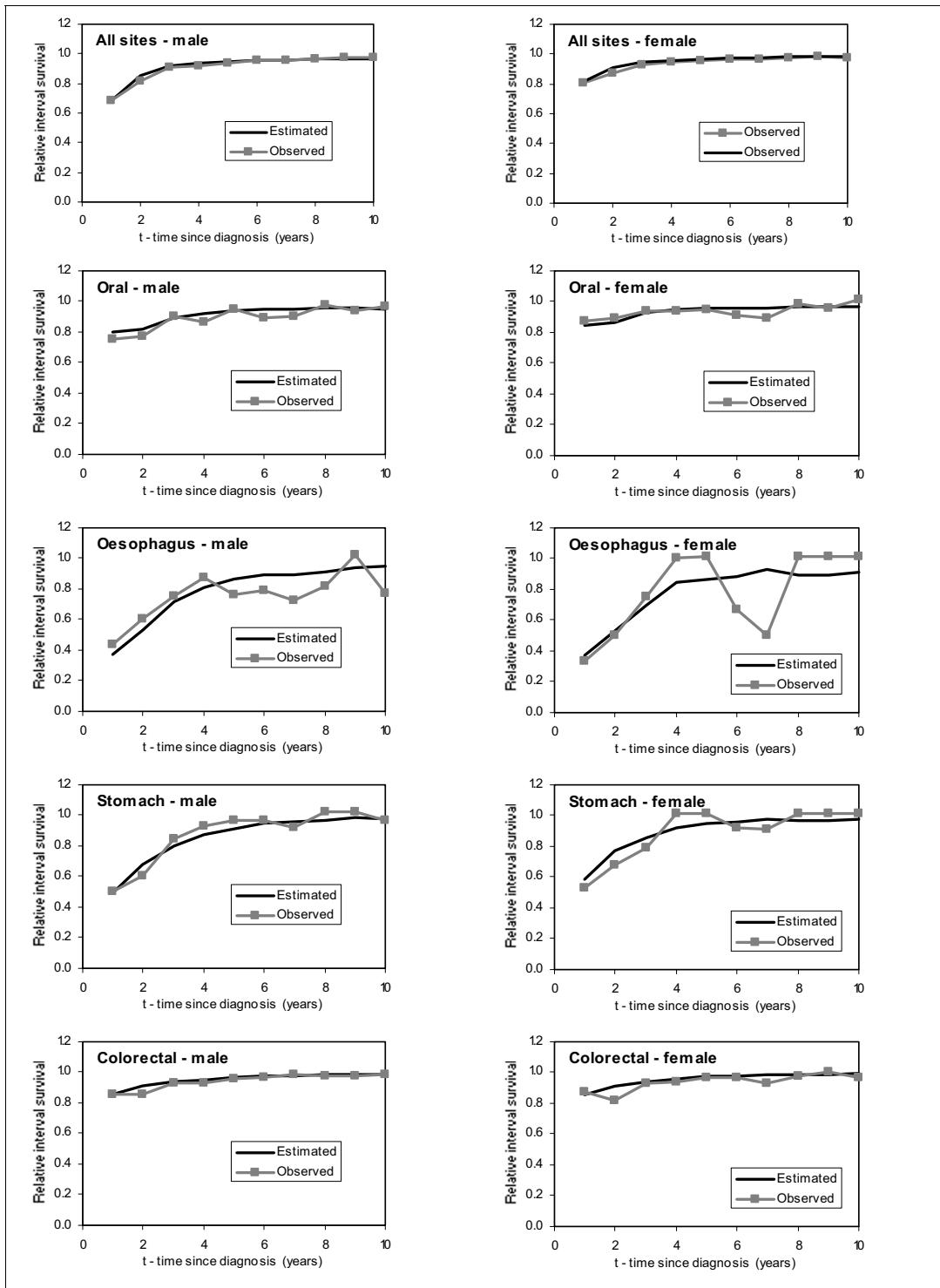
Figure 7
Comparison between predicted and observed relative interval survival for 55–59 year olds for 15 cancer sites, by sex, 1986 (1)
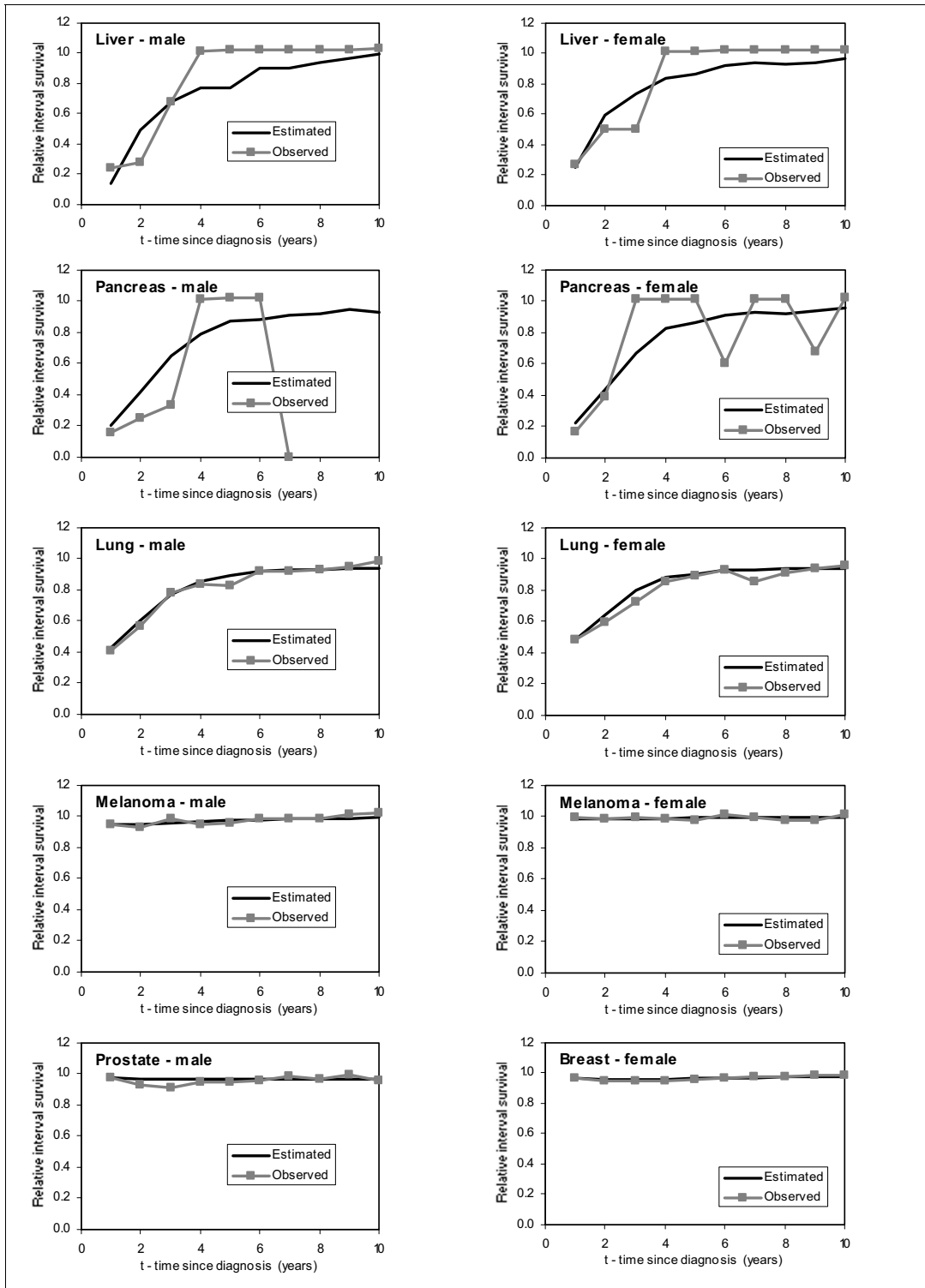
Figure 8
Comparison between predicted and observed relative interval survival for 55–59 year olds for 15 cancer sites, by sex, 1986 (2)
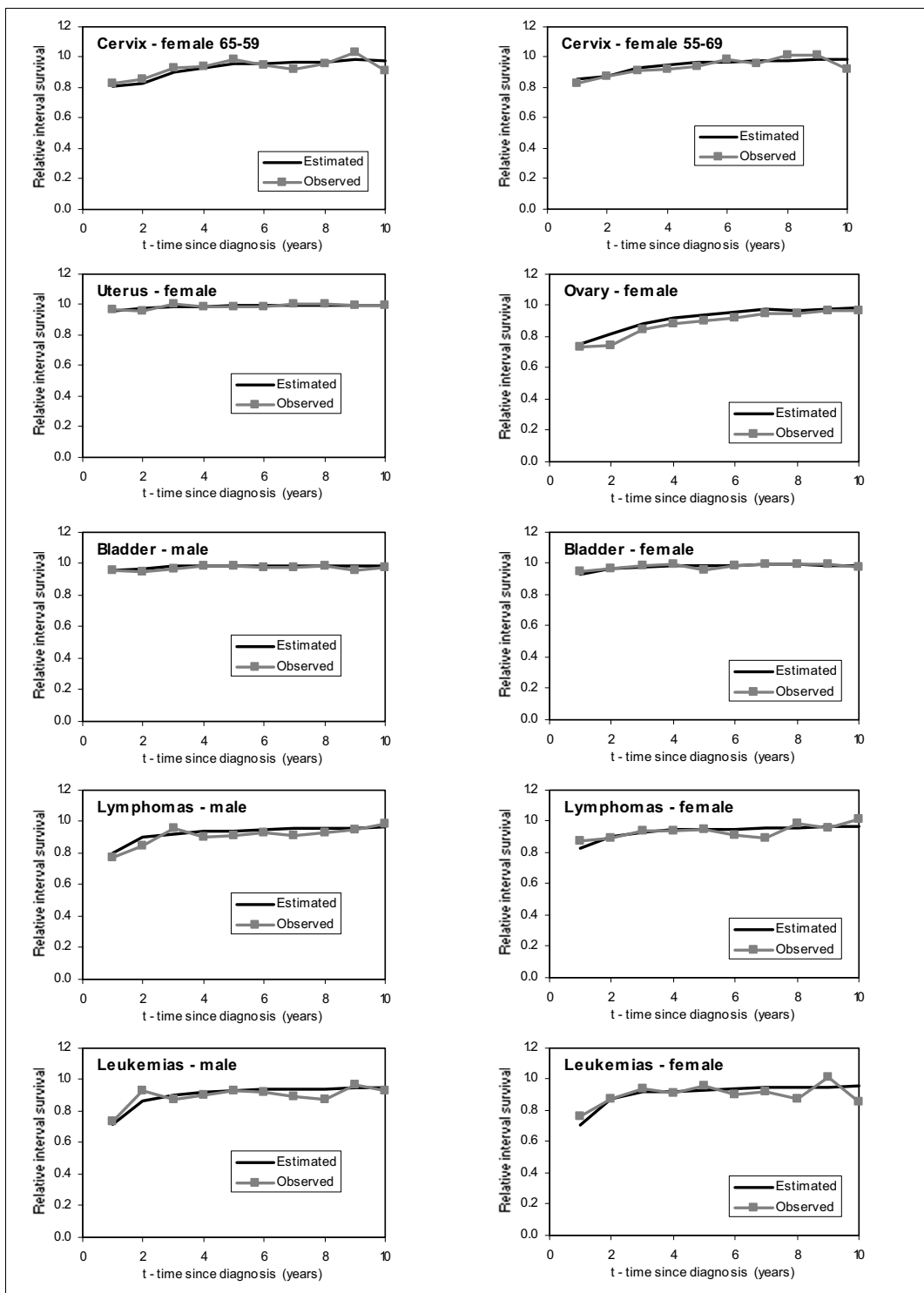
Figure 9
Comparison between predicted and observed relative interval survival for 55–59 year olds for 15 cancer sites, by sex, 1986 (3)
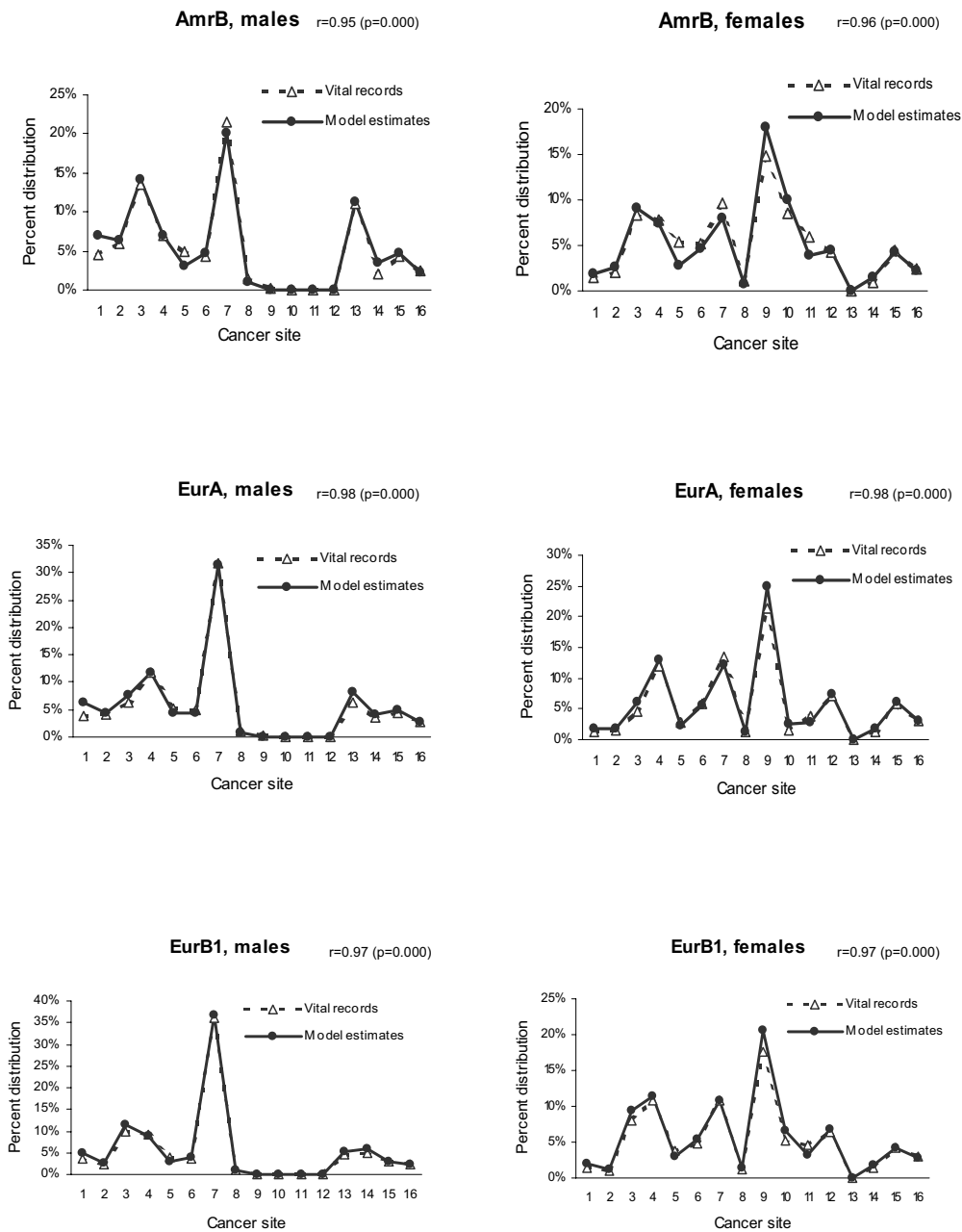
Figure 10
Mortality distribution by site: comparison between model estimates and vital registration data in three sub regions (AmrB, EurA, and EurB1) Cancer site: 1 = mouth and pharynx, 2 = oesophagus, 3 = stomach, 4 = colon and rectum, 5 = liver, 6 = pancreas, 7 = trachea, bronchus and lung, 8 = melanoma of the skin, 9 = breast, 10 = cervix uteri, 11 = corpus uteri, 12 = ovary, 13 = prostate, 14 = bladder, 15 = lymphomas and multiple myeloma, 16 = leukaemia. *r* = Pearson's correlation coefficient when analysed with all data in age groups 45–79.
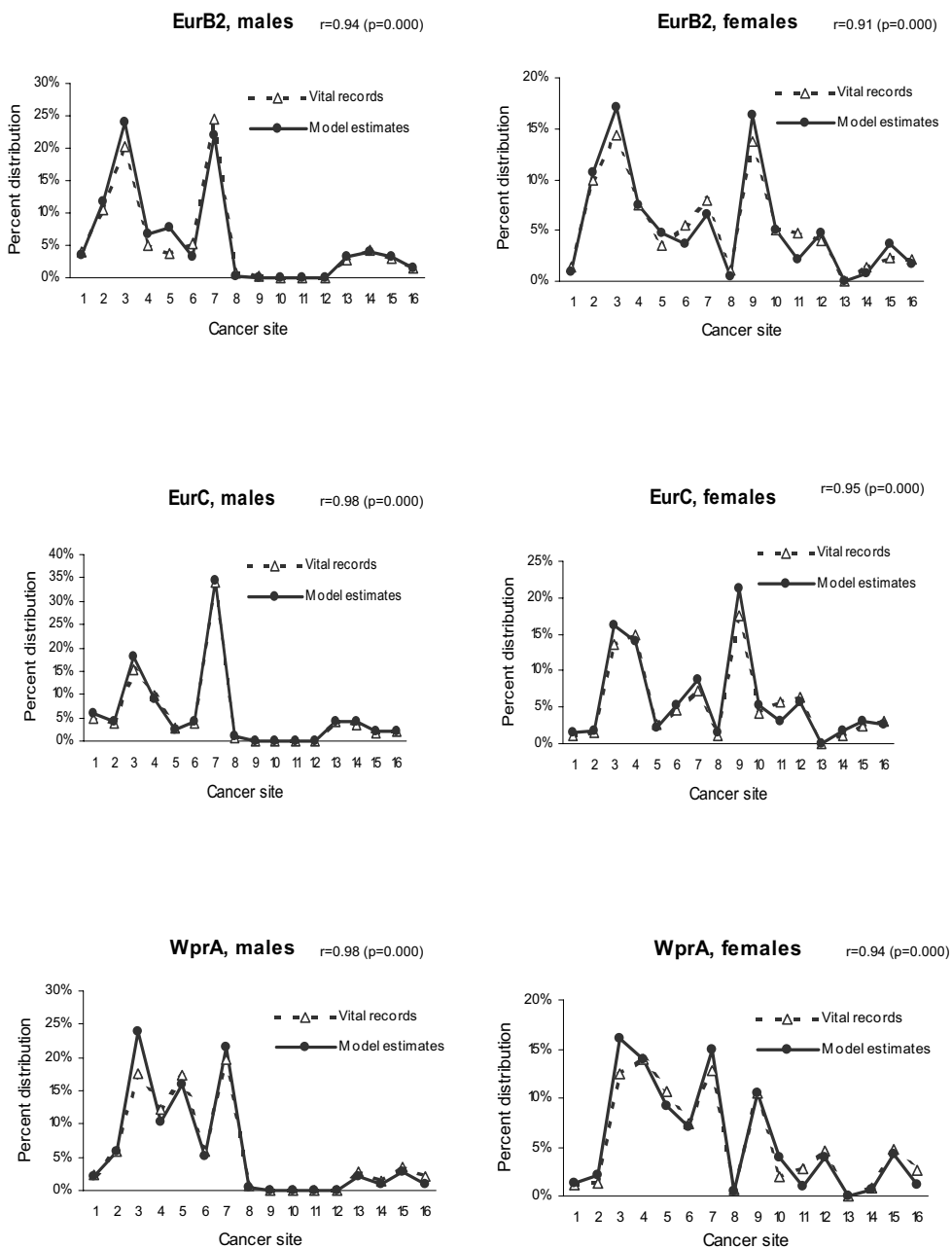
Figure 11
Mortality distribution by site: comparison between model estimates and vital registration data in three sub regions (EurB2, EurC, and WprA) Cancer site: 1 = mouth and pharynx, 2 = oesophagus, 3 = stomach, 4 = colon and rectum, 5 = liver, 6 = pancreas, 7 = trachea, bronchus and lung, 8 = melanoma of the skin, 9 = breast, 10 = cervix uteri, 11 = corpus uteri, 12 = ovary, 13 = prostate, 14 = bladder, 15 = lymphomas and multiple myeloma, 16 = leukaemia. *r* = Pearson's correlation coefficient when analysed with all data in age groups 45–79.
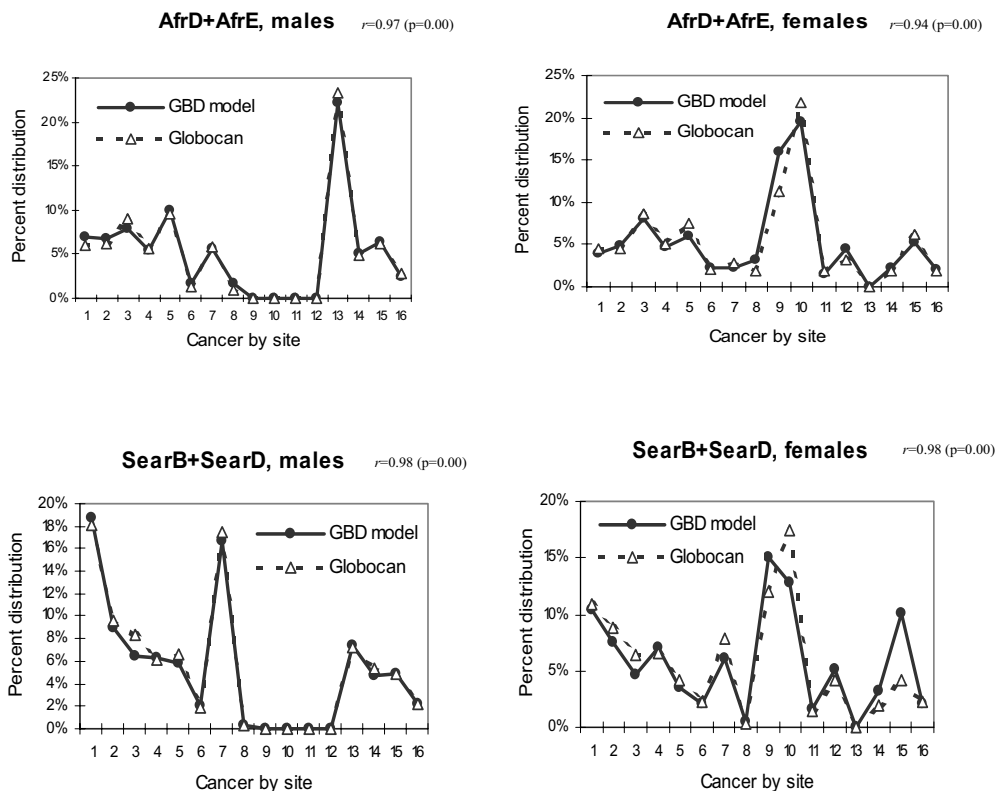
Figure 12
Mortality distribution by site: comparison between model estimates and previous estimates based on cancer registration data in AFRO and SEARO regions Cancer site: 1 = mouth and pharynx, 2 = oesophagus, 3 = stomach, 4 = colon and rectum, 5 = liver, 6 = pancreas, 7 = trachea, bronchus and lung, 8 = melanoma of the skin, 9 = breast, 10 = cervix uteri, 11 = corpus uteri, 12 = ovary, 13 = prostate, 14 = bladder, 15 = lymphomas and multiple myeloma, 16 = leukaemia. *r* = Pearson's correlation coefficient when analysed with all data in age groups 45–79.

plete body of evidence. The second advantage of the proposed approach is that the model is flexible enough to yield the survival estimates of various age, years and period as well as mean duration of time of cancer by site. In addition to mortality and incidence estimates[36], such information is required to estimate the cancer burden in terms of disability-adjusted life years (DALYs) in the future analysis for the GBD 2000[1].

The main limitations for applying this model were the relative lack of region-specific survival data and very few, and probably not always representative, regional cancer incidence data for some developing regions. We assumed that cancer incidence reported by a few countries of one region

or sub region would represent the incidence of the whole area, which may not always be the case.

It is suggested that model-based estimates of cancer mortality in the previous GBD 1990 study did not reflect the actual profile of cancer recorded at the regional registries, in particular of site-specific cancer mortality distribution [3,5,37]. Although population-based estimates from cancer registry data should be incorporated, they may not to be representative of the whole countries they should represent. Such estimates are sometimes restricted to certain geographic areas and also related to the extent of health care and surveillance system. Furthermore, several devel-
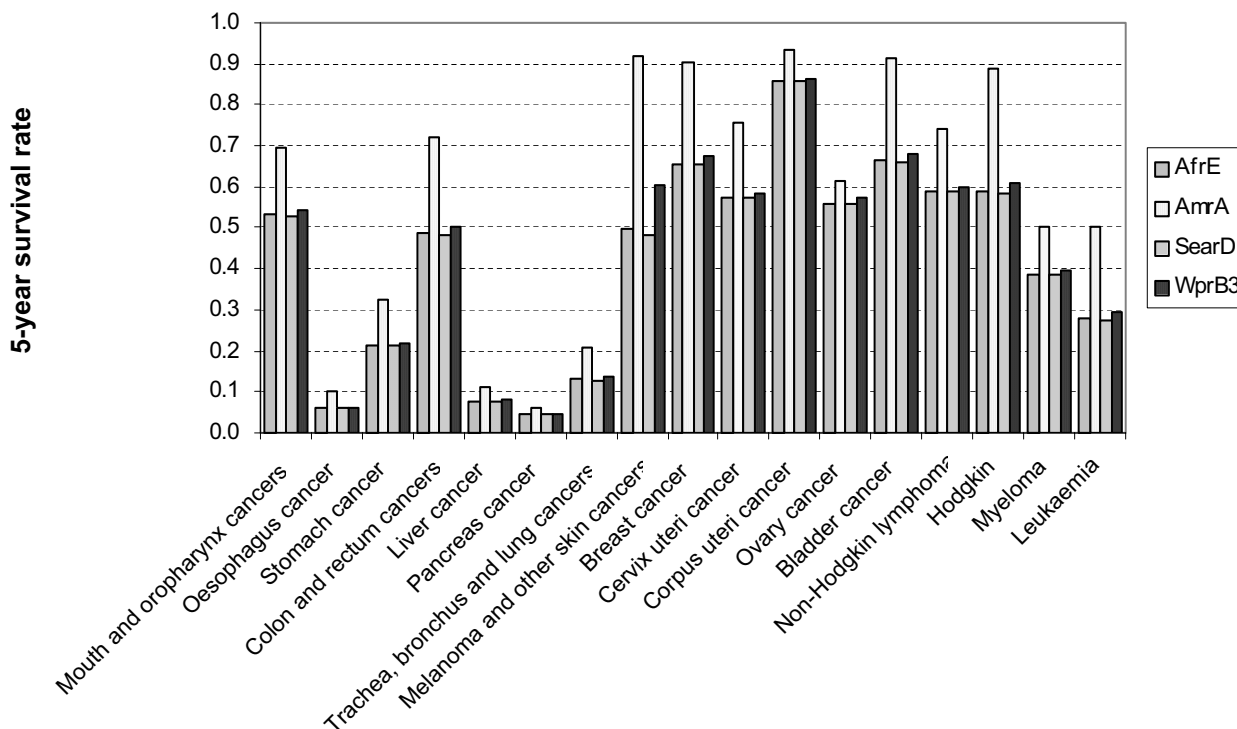
Figure 13
Five-years survival rate by site in four sub regions, females, age 45–54

oping regions of the world were not included in these estimates, and the need to produce model-based estimates would persist.

Contrary to the previous GBD 1990 model, the present survival model specifically developed for the GBD 2000 incorporated all available survival information obtained from registries and corrected for possible bias. The model was used to estimate the distribution of death by site not the actual magnitude of cancer mortality in regions where no or little data on detailed cause of death is available. In fact, the model estimates were quite comparable to the mortality distribution estimated from vital registration records. Furthermore, when for the regions where vital records are not available, our model was consistent with the Globocan 2000 estimates based on the regional registries [6].

Conclusions
The survival model presented offers a new approach to the calculation of the number and distribution of deaths for areas where mortality data are either scarce or unavailable. It can also be applied in areas with good quality data, but where there are small numbers of some site-specific cancers. The model is flexible enough to estimate some of the parameters required to estimate the cancer burden. In our future work, we will attempt to collect further information on survival and incidence from more individual countries in order to improve our estimates, with more precise inputs for the model.

Competing Interests
None declared.
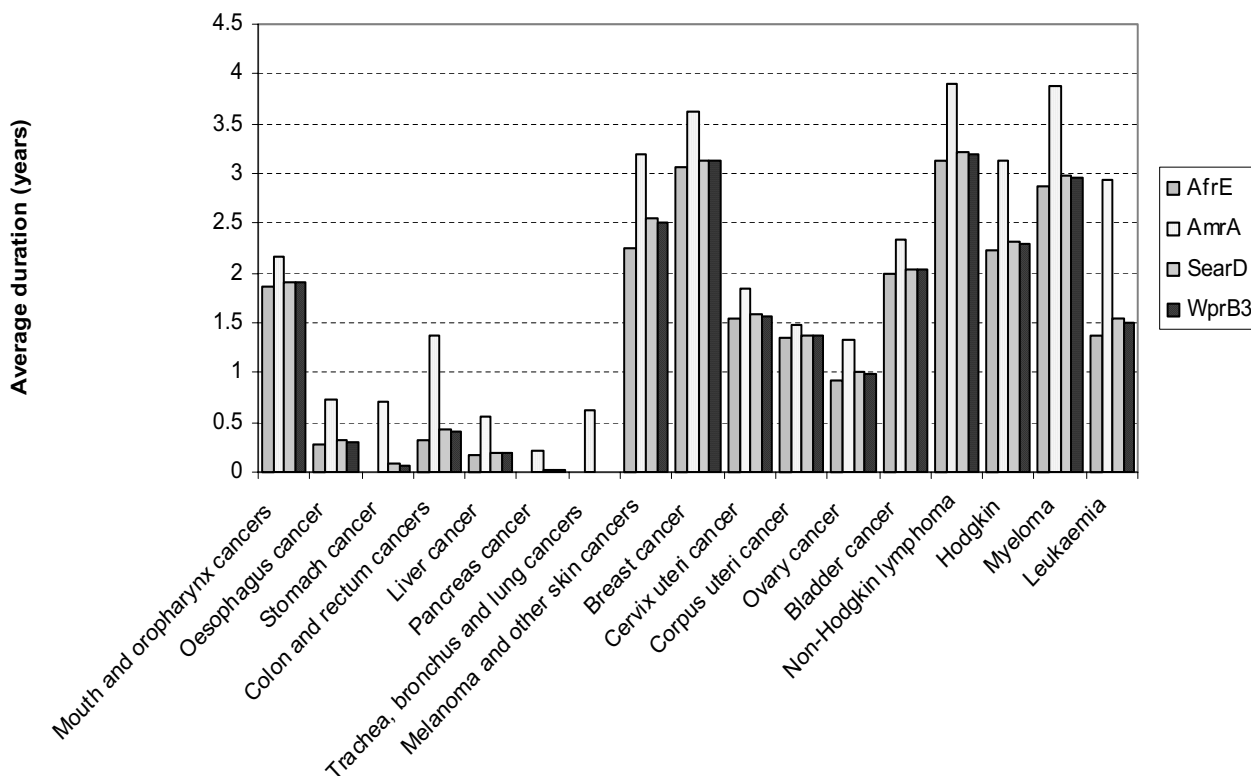
Figure 14
Average duration of patients who die from cancer in four sub regions, females, age 45–54

Author's Contributions

CDM participated in the design of the study, analysed the data, and drafted the initial manuscript. KS participated in the design of the study, performed statistical analyses, and drafted the revised manuscript. CBP implemented the initial version of the survival model, carried out the data compilation, and drafted the initial manuscript. ADL participated in the design of the study and in the mortality analysis. CJLM conceived of the study and participated in its design and coordination. All authors read and approved the final manuscript.

Acknowledgements

References

1. Murray CJL, Lopez AD and Mathers CD **The Global Burden of Disease 2000 project: aims, methods and data sources.** *Geneva, World Health Organization* 2001,
2. Pisani P, Parkin DM and Bray F **Estimates of the worldwide mortality from 25 cancers in 1990.** *Int J Cancer* 1999, **83(1):**18-29
3. Murray CJL and Lopez AD **Mortality by cause for eight regions of the world: global burden of disease study.** *Lancet* 1997, **349:**1269-1276
4. Parkin DM **The global burden of cancer.** *Semin Cancer Biol* 1998, **8:**219-235
5. Parkin DM, Pisani P and Ferlay J **Estimates of the worldwide incidence of 25 major cancers in 1990.** *Int J Cancer* 1999, **80(6):**827-841

6. Ferlay J, Bray F and Pisani P **Globocan 2000: Cancer Incidence, Mortality and Prevalence Worldwide, Version 1.0.** *IARC Cancer Base No. 5. Lyon, IARC Press* 2001,
7. Parkin DM, Whelan SL and Ferlay J **Cancer incidence in five continents.** *IARC Scientific Publications No. 143. Lyon, International Agency for Research on Cancer* 1997,
8. Verdecchia A, Capocaccia R and Egidi V **A method for the estimation of chronic disease morbidity and trends from mortality data.** *Stat Med* 1989, **8:**201-216
9. Lundberg O **Methods of estimating morbidity and prevalence of disablement by use of mortality statistics.** *Acta Psychiatrica Scandinavica* 1973, **49:**324-331
10. Wingo PA, Lndis S and Parker S **Using cancer registry and vital statistics data to estimate the number of new cancer cases and deaths in the United States for the upcoming year.** *Journal of Regulation and Management* 1998, **25:**43-51
11. Ries LAG, Eisner MP and Kosary CL **SEER Cancer Statistics Review, 1973–1999.** *Bethesda, MD, National Cancer Institute* 2002,
12. Sitas F, Bezwoda WR and Levin V **Association between human immunodeficiency virus type 1 infection and cancer in the black population of Johannesburg and Soweto, South Africa.** *Br J Cancer* 1997, **75:**1704-1707
13. Mueller N **Overview of the epidemiology of malignancy in immune deficiency.** *J Acquir Immune Defic Syndr* 1999, **21(Suppl 1):**S5-10
14. Smith C, Lilly S and Mann K **AIDS-related malignancies.** *Ann Med* 1998, **30:**323-344
15. Serraino D **The spectrum of Aids-associated cancers in Africa.** *AIDS* 1999, **13:**2589-2590
16. Sitas E, Pacella-Norman R and Carrara H **The spectrum of HIV-1 related cancers in Sourth Africa.** *Int J Cancer* 2000, **88:**489-492
17. Parkin DM, Garcia-Giannoli H and Raphael M **Non-Hodgkin lymphoma in Uganda: a case-control study.** *AIDS* 2000, **14:**2929-2936
18. Newton R, Grulich A and Sindikubwabo B **Cancer and HIV infection in Rwanda.** *Lancet* 1995, **345:**1378-1379
19. Sitas F, Madhoo J and Wessie J **Cancer in South Africa, 1993–1995.** *Johannesburg, National Cancer Registry of South Africa, South African Institute for Medical Research* 1998,
20. Chokunonga E, Levy LM and Bassett MT **Cancer incidence in the African population of Harare, Zimbabwe: second results from the cancer registry 1993–1995.** *Int J Cancer* 2000, **85:**54-59
21. Wabinga HR, Parkin DM and Wabwire-Mangen F **Cancer in Kampala, Uganda, in 1989–91: changes in incidence in the era of AIDS.** *Int J Cancer* 1993, **54:**26-36
22. Newton R, Ngilimana PJ and Grulich A **Cancer in Rwanda.** *Int J Cancer* 1996, **66:**75-81
23. Bah E, Hall AJ and Inskip HM **The first 2 years of the Gambian National Cancer Registry.** *Br J Cancer* 1990, **62:**647-650
24. Australian Institute of Health and Welfare (AIHW) **Cancer survival in Australia, 2001.** *Part 1: National summary statistics. Canberra, Australian Institute of Health and Welfare* 2001,
25. Martin AA, Galan YH and Rodriguez AJ **The Cuban National Cancer Registry: 1986–1990.** *Eur J Epidemiol* 1998, **14:**287-297
26. Brooks SE, Hanchard B and Wolff C **Age-specific incidence of cancer in Kingston and St. Andrew, Jamaica, 1988–1992.** *West Indian Med J* 1995, **44:**102-105
27. Adib SM, Mufarrij AA and Shamseddine AI **Cancer in Lebanon: an epidemiological review of the American University of Beirut Medical Center Tumor Registry (1983–1994).** *Ann of Epidemiol* 1998, **8:**46-51
28. United Nations **World Population Prospects** *The 1998 revision Volume III: Analytical Report. New York, United Nations* 2000,
29. Surveillance, Epidemiology, and End Results (SEER) **Program Public-Use Data (1973–1999)** *National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April based on the November* 2002,
30. Eisenberg H, Sullivan PD and Connelly RR **Cancer in Connecticut.** *Survival experience. Hartford, Connecticut State Department of Health* 1968,
31. Sankaranarayanan R, Black RJ and Parkin DM **Cancer survival in developing countries.** *IARC Scientific Publications No. Lyon, International Agency for Research on Cancer* 1998,
32. Berrino F, Capocaccia R and Estève J **Survival of cancer patients in Europe: the EUROCARE-2 study.** *IARC Scientific Publications No. 151. Lyon, International Agency for Research on Cancer* 1999,
33. Osaka Prefectural Department of Public Health and Welfare, Osaka Medical Association, Osaka Medical Center for Cancer and Cardiovascular Diseases **Annual Report of Osaka Cancer Registry No. 64-Cancer Incidence and Medical Care in Osaka in 1998 and the Survival in 1994.** *Osaka, Japan, Osaka Prefectural Department of Public Health and Welfare* 2001,
34. Greene WH *Econometric analysis. New York, Prentice Hall* 1997,
35. Sankaranarayanan R, Swminathan R and Black RJ **Global variations in cancer survival.** *Cancer* 1996, **78:**2461-4
36. Shibuya K, Mathers CD and Boschi-Pinto C **Global and regional estimates of cancer mortality and incidence by site: II. Results for the Global Burden of Disease 2000.** *BMC Cancer* 2002, 37
37. Gupta P, Sankaranarayanan R and Ferlay J **Cancer death in India: is the model-based approach valid?** *Bull World Health Organ* 1994, **72:**943-944

Pre-publication history

The pre-publication history for this paper can be accessed here: