Determination of Carcinogenic Characteristics of GI Cancers

Nirosha Rathnayake

Outline

* Introduction
* Main Hypothesis and Assumption
* Time-to-event analysis
* Long and short term goals
* Data preparation
* Computational experiments
* Conclusion and Current Work
* Limitations
* Acknowledgments

Introduction

* Carcinogenesis is the process of cancer development. During this process, normal cells transform into malignant cells.

* Carcinogenic characteristics are age effects, periods effects, cohort effects, population hazard rates, individual hazard rates, and individual survival rates.

Introduction Cont'd

* Majority of cancers occurring in human body is from the gastrointestinal (GI) organ system.
* Therefore, as the test bed, we will use data on the following six GI cancers : pancreas, stomach, liver, gall bladder, colon & rectum, and esophagus.

* On the population level, we will analyze the age effects, the period effects, the cohort effects, and the population hazard rates.
* On the individual level (for individuals susceptible to cancer), we will analyze the individual hazard rates (failure rates), and the individual survival rates (resistance rates).

Main Hypothesis and Assumption

* Population has a dichotomous susceptibility to cancer:

  a small fraction of individuals can experience cancer during their lifetime, while the majority of the population escapes cancer[1].

* We assumed that methods of survival analysis can be adapted for carcinogenic modeling conditioning on susceptibility to cancer.

[1]Mdzinarishvili T, Sherman S, PLoS One, 9(6), 2014.

Time-to-event Analyses

Concepts	Survival Analysis	Carcinogenic Modelling
• Event	• Death of cancer patient	• Diagnosis of cancer.
• Time to event	• # of months passed from cancer diagnosis.	• Age at which cancer was diagnosed.
• Population	• Homogenous (all cancer patients experience the event).	• Dichotomous (only a fraction of population experience the event).
• Data	• Individual (separated survival data for each individual).	• Grouped in age intervals (# of occurrences of cancer in each age interval).
• APC effects	• Not considered	• Considered

Long and Short Term Goals

* **The long term goal** - to develop novel approaches for carcinogenic modeling by adapting mathematical formalism of survival analysis.

* **The short term goal** – To determine and analyze the following carcinogenic characteristics : age, period, cohort (APC) effects, population hazard rates, individual hazard rates (failure rates), and individual survival rates (resistance rates).

Data Preparation

* Surveillance, Epidemiology, and End Results Program (SEER9) databases containing data collected from 1975 until 2009 in nine geographical areas (Atlanta, CT, Detroit, IA, Bay area, Seattle, HI, NM, UT) were utilized. These nine areas we called the “Entire” region.

* Geographical areas of living:

  (1) Eastern region - Atlanta, CT, Detroit, IA;
  (2) Western region - Bay area, Seattle, HI, NM, and UT.

* Grouped data into 7 periods :1975–1979, 1980-1984, 1985-1989, 1990-1994, 1995-1999, 2000-2004, 2004-2009.

* Data for GI cancers were extracted by the SEERStat 8.1.5 software package.

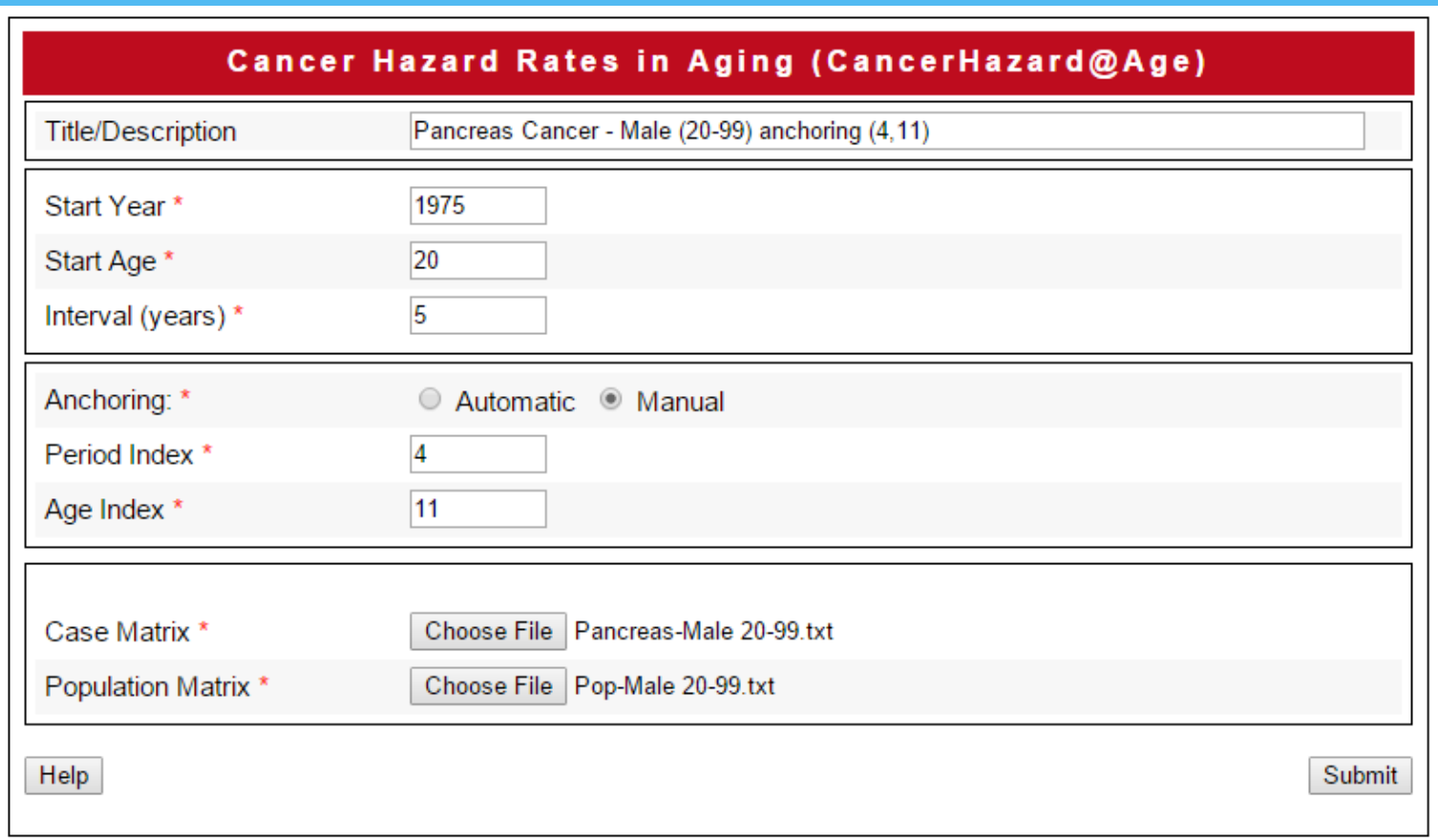
Pancreatic Cancer Cases Diagnosed in Men During 1975-2009 (Entire Area)

Age Interval	1975-1979	1980-1984	1985-1989	1990-1994	1995-1999	2000-2004	2005-2009
20-24	5	4	1	1	3	1	1
25-29	8	3	5	2	7	6	10
30-34	13	14	21	22	17	19	27
35-39	32	46	53	53	54	64	57
40-44	74	89	92	117	127	139	134
45-49	166	170	164	169	251	283	309
50-54	322	320	274	286	399	506	584
55-59	491	516	416	410	470	628	780
60-64	649	659	663	529	585	672	909
65-69	660	671	719	778	676	695	924
70-74	558	609	678	723	716	717	761
75-79	409	436	479	554	539	617	686
80-84	199	232	280	267	342	410	506
85-89	95	103	126	93	119	147	226
90-94	18	23	22	22	29	26	42
95-99	3	2	7	2	3	2	11

Populations of Men Living in the Entire Area During 1975-2009

Age Interval	1975-1979	1980-1984	1985-1989	1990-1994	1995-1999	2000-2004	2005-2009
20-24	4823448	5066913	4740361	4468579	4345198	4794672	4950981
25-29	4557117	5107569	5314620	5037570	4919617	4721690	4971495
30-34	3854083	4647582	5174288	5519031	5373101	5151259	4785983
35-39	3031712	3764909	4578026	5224678	5602105	5283839	5053627
40-44	2655844	2980012	3763370	4646871	5199589	5447194	5177116
45-49	2688338	2573983	2944425	3692897	4529835	5046921	5308329
50-54	2757647	2596258	2495095	2851472	3635290	4432905	4911526
55-59	2517854	2563881	2413126	2360030	2719212	3433799	4234026
60-64	2078337	2251967	2289385	2217312	2184372	2488027	3190864
65-69	1595444	1788748	1963592	2014472	1958011	1927699	2259769
70-74	1134899	1291607	1445855	1621767	1694793	1674177	1690403
75-79	736030	843325	982036	1136567	1302277	1380812	1375331
80-84	440318	469740	542595	650905	786468	920646	1010045
85-89	202239	229444	252456	294440	362880	435645	539711
90-94	69129	78429	86295	100646	124040	148913	184484
95-99	15213	17260	18991	22149	27297	32771	40599

Front-End Interface of "CancerHazard@Age[1]"



[1]Mdzinarishvili T., Sherman, A., Shats O., Sherman S., *Cancer Informatics*, 2014, 13, 197-205. (http://registry.unmc.edu/CHA/index.jsp).

Title/Description	PancMale 20-99		
	Start year	Start age	Time interval
	1975	20	5
	Manual Anchoring	Period Index	Age Index
		4	11

Intercept	SE
-9.375696292	0.0273398
Overall Rate	**SE**
303.8554	6.2341

Table 1 : Period Effects

Index	Period	Period	SE
1	1975-1979	0.1242826	0.0642419
2	1980-1984	0.1075420	0.0419466
3	1985-1989	0.0675000	0.0000000
4	1990-1994	0.0000000	0.0000000
5	1995-1999	-0.0153442	0.0415522
6	2000-2004	0.0220117	0.0636028
7	2005-2009	0.0962116	0.0869532

Table 2: Age Effects

Index	Age	Effect	SE
1	20-24	-6.6550497	0.3746646
2	25-29	-5.9808007	0.2954005
3	30-34	-4.8076189	0.2226752
4	35-39	-3.7023024	0.1824736
5	40-44	-2.8508066	0.1531122
6	45-49	-2.0705202	0.1259754
7	50-54	-1.3814877	0.1006308
8	55-59	-0.9048726	0.0764068
9	60-64	-0.4839129	0.0530060
10	65-69	-0.1752762	0.0328114
11	70-74	0.0000000	0.0000000
12	75-79	0.0531236	0.0340469
13	80-84	0.0142365	0.0563864
14	85-89	-0.1635389	0.0841632
15	90-94	-0.7069140	0.1303460
16	95-99	-0.8516781	0.2447772

Table 3 : Cohort Effects

Index	Cohort	Effect	SE
1	1880	-0.0376395	0.6873234
2	1885	0.0307182	0.3070006
3	1890	0.1866968	0.1790180
4	1895	-0.0048238	0.1371230
5	1900	0.0981851	0.1048672
6	1905	0.0370353	0.0789823
7	1910	0.0014469	0.0556728
8	1915	0.0213285	0.0354991
9	1920	0.0000000	0.0000000
10	1925	0.0311958	0.0342838
11	1930	-0.0189861	0.0546381
12	1935	0.0013131	0.0770987
13	1940	0.0158028	0.1004602
14	1945	0.0017681	0.1244081
15	1950	0.0130291	0.1490555
16	1955	0.0168173	0.1747597
17	1960	-0.0020914	0.2021856
18	1965	-0.0080460	0.2347224
19	1970	0.0100347	0.2769506
20	1975	0.3399706	0.3372340
21	1980	0.3962086	0.4749086
22	1985	-1.0899998	1.1612424

Table 4 : Population Hazard Rates			
Index	Age	Pop.Rate$(h_u(t_i))$	SE
1	20-24	0.0109128	0.0040995
2	25-29	0.0214170	0.0063536
3	30-34	0.0692252	0.0155305
4	35-39	0.2090727	0.0385761
5	40-44	0.4898885	0.0761943
6	45-49	1.0689843	0.1378006
7	50-54	2.1291897	0.2220288
8	55-59	3.4293015	0.2782907
9	60-64	5.2242762	0.3115831
10	65-69	7.1132054	0.3037973
11	70-74	8.4759198	0.2317296
12	75-79	8.9383660	0.3902962
13	80-84	8.5974506	0.5387582
14	85-89	7.1971869	0.6368962
15	90-94	4.1800164	0.5567046
16	95-99	3.6166606	0.8907810

Table 5 : Individual Hazard Rates (Failure Rates)			
Index	Age	Ind. Rate	SE
1	20-24	0.0000359	0.0000135
2	25-29	0.0000705	0.0000210
3	30-34	0.0002281	0.0000515
4	35-39	0.0006904	0.0001286
5	40-44	0.0016271	0.0002564
6	45-49	0.0035971	0.0004728
7	50-54	0.0073627	0.0007915
8	55-59	0.0124571	0.0010655
9	60-64	0.0205960	0.0013633
10	65-69	0.0319248	0.0017030
11	70-74	0.0461053	0.0021631
12	75-79	0.0637078	0.0041484
13	80-84	0.0891270	0.0080591
14	85-89	0.1263188	0.0165228
15	90-94	0.1464958	0.0308972
16	95-99	0.4000000	0.1393279

0.9983289
0.9983289

Individual Survival Rates

* Individual survival rates (resistance rates) were calculated based on the population hazard rates as follows[3]:

$$s(t_i) = \frac{H_o - H_U(t_i)}{H_0}$$

Where, $H_U(t_i) = \Delta \sum_1^i h_u(t_i)$ - cumulative hazard function,
H_0 - overall cumulative hazard value, and
t_i – the i^{th} age interval.

Ex: $H_o = 303.8554$, for 3rd age interval: $H_U(t_3) = 5 \times \sum_1^3 h_u(t_i) = 0.507775$

$$s(t_i) = \frac{303.8554 - 0.507775}{303.8554} = 0.9983289$$

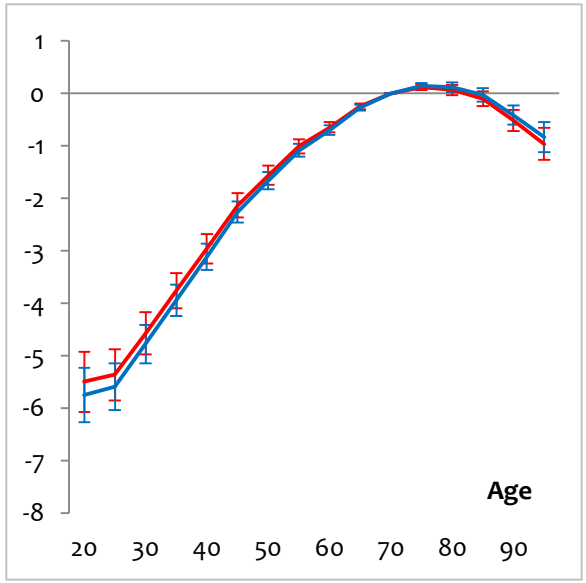
[3]Mdzinarishvili T., Sherman S. Cancer Informatics, 2013, 12, 67-81.

Computational Experiments

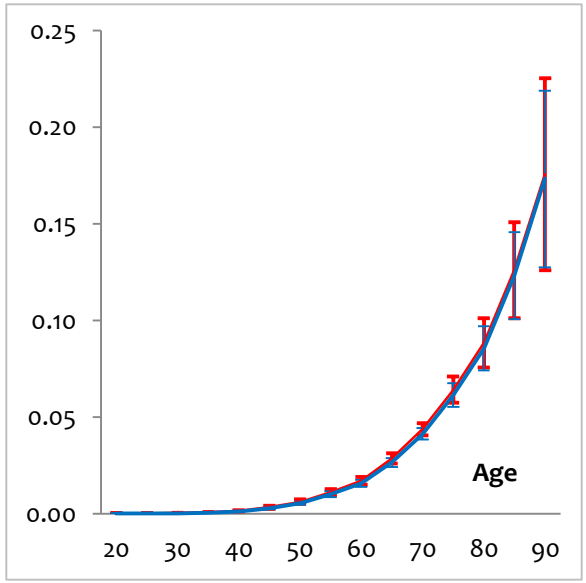
* Two series of computational experiments for six GI cancers were performed to determine:

1. How the age effects, period effects, and cohort effects, as well as the population hazard rates, failure rates, and resistance rates depend on time period, when these cancers were diagnosed?
2. How the aforementioned characteristics depend on the geographic areas, where these cancers were diagnosed?

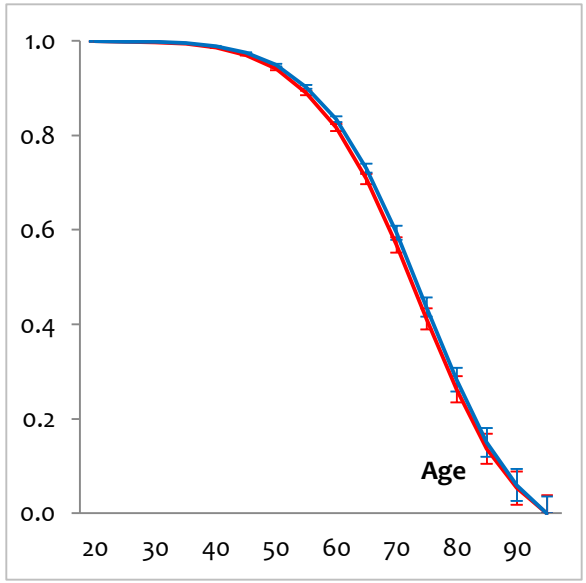
Carcinogenic Characteristics of Pancreatic Cancer in Men that are Invariant to Time Period of Diagnosis



Age effects



Individual hazard rates

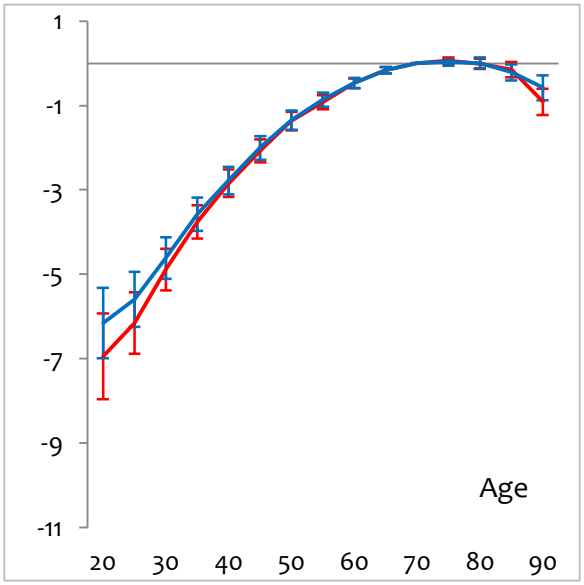


Individual survival rates

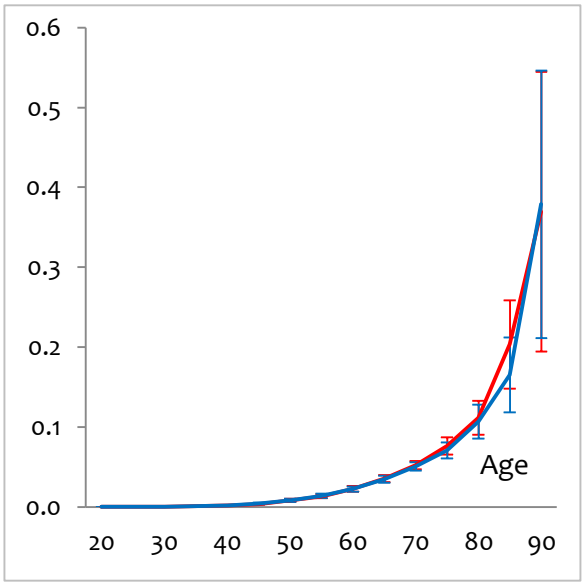
anchored to 1990 -1994 time period

anchored to 2005 -2009 time period

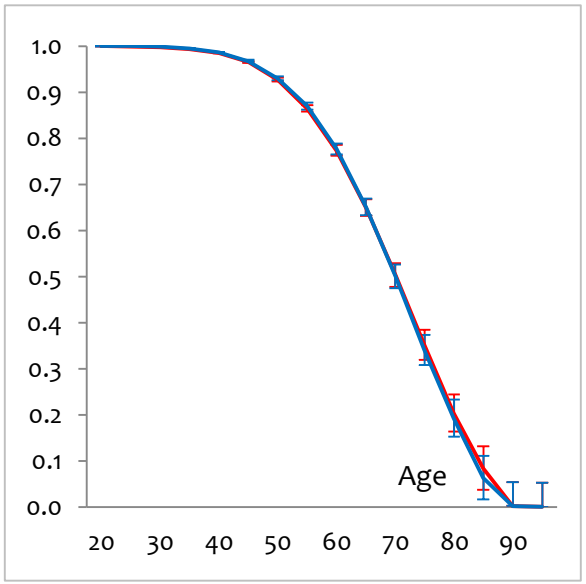
Carcinogenic Characteristics of Pancreatic Cancer in Men that are Invariant to Place of Living



Age effects



Individual hazard rates



Individual survival rates

Eastern region (Atlanta, CT, Detroit and IA)

Western region (Bay area, Seattle, HI, NM, and UT)

Conclusion and Current Work

* Conclusion :
  * The age effects, the failure rates (individual hazard rates), and the resistance rates (individual survival rates) in aging stratified by gender are nearly independent on geographical area and the time periods of cancer diagnosis.
* Current Work:
  * Studying patterns changes of aforementioned characteristics of six GI cancers depend of gender.
  * Developing novel approaches for carcinogenic modelling based survival analysis.
  * A manuscript presenting the obtained results is in preparation.

Limitations of Our Study

* Our protocol does not fully account for heterogeneity of data. It stratifies data only by gender. Further stratification (based on race, subtypes of cancer and staging at diagnosis, etc.) is required to get more homogeneous data

Acknowledgements

* **Simon Sherman,** Ph.D., professor, Eppley Cancer Institute, UNMC.
* **Gleb Haynatzki,** PhD., DSc, Professor, College of Public Health,UNMC
* **Tengiz Mdzinarishvili**, Ph.D., D.S., former Research Assistant Professor at Dr. Sherman's laboratory.
* **Oleg Shats**, M.S., IT manager at Dr. Sherman's laboratory.

QUESTIONS?

APC Effects, Population and Individual Hazard Rates

* APC effects, population hazard rates, and individual (failure) hazard rates were calculated using the "*CancerHazard@Age*" web tool.[1]

* The aforementioned carcinogenic characteristics were obtained by using two anchors:
    1. 70-74 age interval and 1990-1994 period ($[i,j] = [11,4]$),
    2. 70-74 age interval and 2005-2009 period ($[i,j] = [11,7]$).

APC formula:

$$k = j - i + n \ (i = 1, \dots, 16;\ j = 1, \dots, 7, n = 16)$$

k – index for the birth cohorts; i – index for the five-year-long age intervals; j – index for the five-year-long time periods.

Ex: For 4th periods index and 11th age index : $k = j - i + n = 9$
Baseline A, P, C, will be 11, 4, and 9.

[1]Mdzinarishvili T., Sherman, A., Shats O., Sherman S., *Cancer Informatics*, 2014, 13, 197-205. (http://registry.unmc.edu/CHA/index.jsp).