ASSESSMENT OF CHANGE IN SOCIOECONOMIC STATUS EFFECT ON CANCER SURVIVAL IN ONTARIO FROM 1993 – 2009

by

Andrew Dabbikeh

A report submitted to the Department of Public Health Sciences

In conformity with the requirements for

the degree of Master of Science

Queen's University

Kingston, Ontario, Canada

(August, 2015)

Supervisors: Dr. Paul Peng and Ms. Jina Zhang-Salomons

Abstract

Background: It is known that cancer survival is dependent on a person's socioeconomic status (SES). However, little is known about whether this association has evolved over time.

Objectives: This study will assess whether the difference in cancer survival between the richest and poorest SES groups changes between 1993 and 2009 in Ontario, Canada.

Methods: Between 1993 and 2009, 920,334 cancer cases were identified from the Ontario Cancer Registry. Neighborhood median household income from the Canadian census was linked to the registry. 5-year cancer-specific (CSS) survival was calculated by SES quintiles and year of diagnosis. A Fine-Gray model was used to assess the interaction between SES and year of diagnosis from 1993 to 2009, controlling for age and sex while treating non-cancer deaths as competing events.

Results: It was found that survival has improved during the study period, varying by SES. A strong and significant interaction between SES and year of diagnosis was observed for all cancer cases. The hazard decreased by 3.1% per year of diagnosis in the richest SES group, and by 1.2% per year of diagnosis in the poorest SES group. There was a strong and significant interaction effect among breast and colorectal cancer cases. A weak, yet significant, interaction was found for lung cancer as well as head and neck cancer.

Conclusions: Cancer survival in Ontario has improved for the richest communities at a faster rate relative to the poorer communities. Future research should examine the possible causes, such as screening procedures, treatment adherence, and/or economic disparity.

Acknowledgements

Firstly, I would like to thank my supervisors Dr. Paul Peng and Ms. Jina Zhang-Salomons for offering me such a great opportunity to study in the epidemiology specializing in biostatistics program for my Master of Science. I feel thankful for their ongoing support and instrumental ideas throughout my courses as well as my practicum project. Their guidance improved my understanding of statistics, especially when applied to the wide field of epidemiology, as well as stimulated my interest in learning statistical techniques.

Also, I greatly appreciate the help from faculty members in both the Department of Public Health and Department of Mathematics and Statistics. They provided me guidance and experience in the field of statistics and epidemiology.

I am very thankful of my biostatistics colleagues – Laura, Katherine, Fiona, and Jessica. I will always remember the help we provided for one another during the course of the year. I'm most thankful of my family and friends, who supported and encouraged me throughout my time at Queen's.

In the end, I would like to thank Cancer Care Ontario for the financial support and computation facilities.

Table of Contents

ABSTRACT……**i**

ACKNOWLEDGEMENTS……**ii**

TABLE OF CONTENTS……**iii**

LIST OF FIGURES……**iv**

LIST OF TABLES……**v**

1. INTRODUCTION AND LITERATURE REVIEW……**1**
2. METHODS……**2**

2.1. SOURCE OF DATA……**2**

2.2. STATISTICAL ANALYSIS……**4**

3. RESULTS……**6**

3.1. STUDY POPULATION PROFILE……**6**

3.2. SURVIVAL BY YEAR OF DIAGNOSIS……**8**

3.3. SURVIVAL BY SES……**9**

3.4. SES EFFECT BY YEAR OF DIAGNOSIS……**11**

4. DISCUSSION……**19**
5. SUMMARY……**21**
6. REFERENCES……**21**
7. APPENDIX A – OVERALL SURVIVAL BY YEAR OF DIAGNOSIS……**24**
8. APPENDIX B – SES TREND AND OVERALL SURVIVAL……**25**
9. APPENDIX C – FINE-GRAY MODEL……**27**
10. APPENDIX D – MISSING DATA……**28**
11. APPENDIX E – SCHOENFELD RESIDUALS……**30**
12. APPENDIX F – SAS CODE: CREATING AND ORGANIZING DATASET……**38**
13. APPENDIX G – SAS CODE: ANALYZING DATASET……**46**

List of Figures

Figure 1. The 5-year cancer-specific survival, of all cancer cases in addition to breast, lung, head and neck, and colorectal cancer cases in Ontario from 1993 to 2006. The trend line was included for each cancer group.……..……..……..**9**

Figure 2. The 5-year cancer-specific survival, including 95% confidence intervals, of all cancer cases, as well as deaths due to breast, lung, head and neck, and colorectal cancer cases in Ontario from 1993 to 2006, separated by SES quintiles…….……**10**

Figure 3. The 5-year cancer-specific survival of all cancer cases in Ontario from 1993 to 2006, separated by SES and year of diagnosis.……..…….……**12**

Figure 4. The 5-year cancer-specific survival of breast cancer (a), lung cancer (b), colorectal cancer (c), and head & neck cancer (d) in Ontario from 1993 to 2006, separated by SES and year of diagnosis.…….…….…….……..……...……...**15**

Figure 5. The 3-year overall survival, of all cancer cases in addition to breast, lung, head and neck, and colorectal cancer cases in Ontario from 1993 to 2009. These values were calculated from the Kaplan-Meier estimator. The trend line was included for each cancer group.……..……...……**25**

Figure 6. The 3-year overall survival and cancer-specific survival, including their 95% confidence intervals, for all cancer cases in Ontario from 1993 to 2009, separated by SES quintiles…….**26**

Figure 7. The percentage of missing cases in each year of diagnosis in Ontario from 1993 to 2009……**29**

Figure 8. The 3-year overall survival, including 95% confidence intervals, for all cancer cases reporting median household income data (Available) and cases without a reported median household income (Missing), from 1993 – 2009. Calculated using the Kaplan-Meier estimator……..……...……**30**

Figure 9-22. Graphic analysis of proportional hazards assumption using Schoenfeld residuals (by each covariate)……...…….……**31 – 37**

List of Tables

Table 1. Description of Cancer Cases in Ontario from 1993 – 2009……...……**7**

Table 2. The hazard ratio per year of diagnosis, separated by SES, for cancer cases in Ontario diagnosed between 1993 and 2009. The hazard ratios and their 95% confidence intervals were calculated from the Fine-Gray model……..……..**16**

Table 3. Summary of Fine-Gray model for all cancer cases in addition to breast, lung, head and neck, and colorectal cancer cases in Ontario……...……..……**17**

1. INTRODUCTION AND LITERATURE REVIEW

For decades, it has been known that cancer survival is associated with socioeconomic status (SES) in developed countries such as the United States, United Kingdom, and Canada[1,13,14]. For example, in a 1997 Canadian study, the authors found that SES, represented as median household income, was strongly and significantly associated with cancer survival among cancers of the head and neck region, breast, bladder, and esophagus. The authors also found weaker, yet significant, associations among cancers of the lung and rectum. They did not, however, find significant associations among cancers of the pancreas, stomach, ovaries, and colon[2].

While the public health care system in Canada is believed to be advantageous for the poorer groups relative to the United States, at least prior to the Affordable Care Act in 2010, a discrepancy in cancer survival between SES groups still exists in Ontario[2]. This discrepancy has been shown to be smaller in Ontario than in the United States, due to better cancer survival in the poorest communities and worse survival in the wealthier communities[13]. Furthermore, income inequality has been increasing in Canada over the past 20 years, even faster than in the United States.

Although the association between SES and cancer survival has been reported in the last two decades, few studies have assessed whether this association has changed over time[1,2,3,4]. Only one Canadian study looked into this time trend in head and neck cancer cases across Canada from 1992 to 2005[5]. The authors found a significant increase in the magnitude of the difference in cancer survival between the richest and poorest SES quintiles for oropharynx cancer only. On the other hand, no significant changes were found among cancers in other regions of the head and neck. A prospective cohort study in Britain looked

into the same time trend among 7,489 men for 35 years[6]. The authors separated the men by whether the men were from manual or non-manual social classes, a similar concept to SES. What the authors observed was no change in the difference in cancer survival between the manual and non-manual groups. They concluded that SES inequality in cancer survival in Britain had remained unchanged within the two groups during the 35-year follow-up period.

An American study examined whether the disparity in cancer survival by insurance status, a closely-related concept, changed between 1999 and 2004[7]. The authors found that survival improved between 1999 and 2004 for privately-insured patients with breast, lung, colorectal, and non-Hodgkin lymphoma. They also found that among Medicaid and uninsured patients, only the survival of the non-Hodgkin lymphoma patients insured by Medicaid improved during the study period.

The objective of this study is to examine whether the association between SES and cancer survival has changed between 1993 and 2009 in Ontario, Canada. A null hypothesis that no change in the difference in cancer survival improvement by SES group was assumed before the study began.

2. METHODS

2.1. Source of Data

This is a population-based retrospective study using data from the Ontario Cancer Registry (OCR) and Statistics Canada. Cancer data was collected from the OCR, which contains all cancer cases diagnosed in Ontario since 1964. For this study, 920,334 cases diagnosed between 1993 and 2009 were identified. The OCR file provides the ICD-O-3

(International Classification of Diseases for Oncology, 3rd edition) codes for each cancer case, which are defined over the study period from 1993 to 2009. For this study, the following sites were examined – breast, lung, head and neck region, colorectal, and all cancers combined. These sites were chosen due to their high prevalence during the study period. The OCR also provided ICD9 and ICD10 codes for cause of death from 1993 to December 31, 2011, while the date of death is complete up to December 31, 2013. The OCR also provides the patients' demographic information such as age, sex, and postal code at diagnosis.

Median household income was collected from the 1996, 2001, and 2006 censuses at the dissemination area or enumeration area (DA/EA) level, from Statistics Canada. The DA/EAs were grouped into 5 quintiles based on their median household income, with the 5^{th} quintile representing where the wealthiest 20% in Ontario resided and the 1^{st} quintile representing where the poorest 20% in Ontario resided.

The cancer and income data were merged into one dataset by the patient's postal code, assigning each patient the appropriate SES quintile based on the median household income of the DA/EA in which they resided at the time of diagnosis. This was done by utilizing the Postal Code Conversion File (PCCF+), which provided the postal code of greatest fit in each DA/EA. The median household income from the 1996 census was used to represent SES for all cancer cases from 1993 to 1997. The 2001 census represented all cancer cases from 1998 to 2002. The 2006 census represented all cancer cases from 2003 to 2009. These years were chosen due to the fact that they represented 2 years prior and 2 years after the recorded median household income of the DA/EA.

2.2. Statistical Analysis

Cancer-specific survival was calculated for each of the groups defined by SES quintiles, by year of diagnosis, and by both SES and year of diagnosis. Cancer-specific survival was chosen over overall survival because cancer-specific survival more directly reflects the outcome of cancer care. The five-year cancer-specific survival is calculated as 1 minus the cumulative incidence function for death from any cancer at five years of follow-up from the date of diagnosis, which takes the competing deaths due to other causes into account. Cases diagnosed between 1993 and 2006 were included. We excluded the cases diagnosed between 2007 and 2009 because of the lack of follow-up for cases diagnosed after 2006 and the lack of cause of death information after 2011. The number of follow-up months was calculated from the date of the patient's diagnosis to the date of death from cancer, date of death from other causes, or until the last date of follow-up, if alive by the end of the study period. Censoring occurred if the patient survived past December 31, 2011, or until follow-up ended. Patients who died from causes other than cancer were treated as competing events.

A Fine-Gray subdistribution hazards regression was used to run a competing risks analysis to test the interaction between SES and year of diagnosis, controlling for age and sex, on the subdistribution of time to death of the cancer[8,9]. The subdistribution hazards are the probabilities in which the patient will fail from the event of interest (cancer), where patients who fail from a cause other than the event of interest remain in the risk set, the set of patients at risk of the event at time *t*. For patients who experience a competing event, the subdistribution hazards accounts for the patient no longer having

any chance of a failure from the event of interest as a result of the competing event. Since the Kaplan-Meier estimate of survival becomes biased if competing events are present, it can no longer be assumed that the patient will experience the event of interest if the follow-up period is long enough. As such, the cumulative incidence function, the marginal failure subdistribution of a given cause, was calculated to estimate the five-year cancer-specific survival.

The Fine-Gray regression models covariate effects on the cumulative incidence of the event of interest in a similar fashion to the Cox proportionality hazards model, assuming that the covariates have proportional effects on the baseline subdistribution hazard. The regression coefficients in the Fine-Gray model are based off of a modified risk set, where patients that experience a competing event are retained after the event. The patients who are retained after the competing event are steadily down-weighted according to the conditional probability of being under follow-up had the competing event not occurred, reflecting the increasing likelihood of censoring[8].

In the studies examining the time trend, the authors mostly relied on overall survival and the Cox proportional hazards model[5,7,13]. While this method is appropriate for overall survival, it is limited for cancer-specific survival due to the fact that not all cancer patients die of cancer. Competing events, like car accidents or heart attacks, prevent the occurrence of deaths related to cancer. Furthermore, with cause of death information available, it was more appropriate to apply cancer-specific survival as well as the Fine-Gray model when examining the interaction between SES and year of diagnosis. Statistical details on the Fine-Gray model can be found in Appendix C.

The results were considered significant at the 0.05 level, and all tests of statistical significance were two-sided. The statistical analysis for this study was performed using SAS 9.4 (SAS Institute Inc., Cary, NC, USA).

3. RESULTS

3.1. Study Population Profile

Table 1 presents summary statistics of the study population. Patients aged 50 years or younger represented roughly 15% of all cancer cases. Meanwhile, patients aged 50-79 made up approximately 69% of all cancer cases. Males represented a larger proportion of cancer cases than females. The number of incident cancer cases grew from 44,165 in 1993 to 65,522 in 2009. The percentage of cancer cases in Ontario was 15.0% in the richest SES quintile and 22.9% in the poorest SES quintile. Breast cancer represented the highest percentage of all cancer cases in the study, followed by lung, colorectal, and head and neck cancers, respectively.

Table 1. Description of Cancer Cases in Ontario from 1993 – 2009

Variables	Groups	# of Cases	Percentage	Median Income ($)
Age	<50	136515	14.8%	
	50-59	154531	16.8%	
	60-69	234047	25.4%	
	70-79	249900	27.2%	
	80+	145341	15.8%	
Sex	Males	477336	51.90%	
	Females	442998	48.10%	
Year of Diagnosis	1993	44165	4.80%	
	1994	44847	4.87%	
	1995	44738	4.86%	
	1996	46082	5.01%	
	1997	47804	5.19%	
	1998	49403	5.37%	
	1999	51164	5.56%	
	2000	52811	5.74%	
	2001	54559	5.93%	
	2002	55452	6.03%	
	2003	56098	6.10%	
	2004	58477	6.35%	
	2005	59971	6.52%	
	2006	61439	6.68%	
	2007	63909	6.94%	
	2008	63893	6.94%	
	2009	65522	7.12%	
SES	1 (Poorest)	210539	22.88%	40,602.61
	2	197432	21.45%	54,640.21
	3	180032	19.56%	67,240.18
	4	157701	17.14%	81,819.97
	5 (Richest)	137659	14.96%	106,019.90
	Missing	36971	4.02%	
Site	Breast	124221	13.50%	
	Lung	122889	13.35%	
	Head & Neck	30695	3.34%	
	Colorectal	122183	13.28%	
	Others	520346	56.53%	
	Total	920334		

3.2. Survival by Year of Diagnosis

Figure 1 shows the 5-year cancer-specific survival rates by year of diagnosis for 4 major sites of cancer and for all cancers combined; the actual survival rates were represented by the markers, with the lines illustrating a linear trend of the markers. Major improvements in cancer-specific survival were found in breast and colorectal cancer cases as well as all cancer cases. The 5-year cancer-specific survival for all cancer cases diagnosed in Ontario improved from 57.8% (95% CI: 57.3 – 58.2) in 1993 to 64.1% (95% CI: 63.8 – 64.5) in 2008. The 5-year cancer-specific survival of breast cancer cases improved from 79.6% (95% CI: 78.6 – 80.6) in 1993 to 85.2% (95% CI: 84.4 – 86.0) in 2006. For colorectal cancer cases, 5-year cancer-specific survival improved from 54.3% (95% CI: 53.0 – 55.6) in 1993 to 61.8% (95% CI: 60.8 – 62.9) in 2006. Lung cancer cases in Ontario showed little improvement in 5-year cancer-specific survival, from 21.6% (95% CI: 20.6 – 22.6) in 1993 to 22.3% (95% CI: 21.4 – 23.3) in 2006. Head and neck cancer cases also showed little improvement in 5-year cancer-specific survival, from 63.3% (95% CI: 60.9 – 65.7) in 1993 to 66.7% (95% CI: 64.6 – 68.8) in 2006.

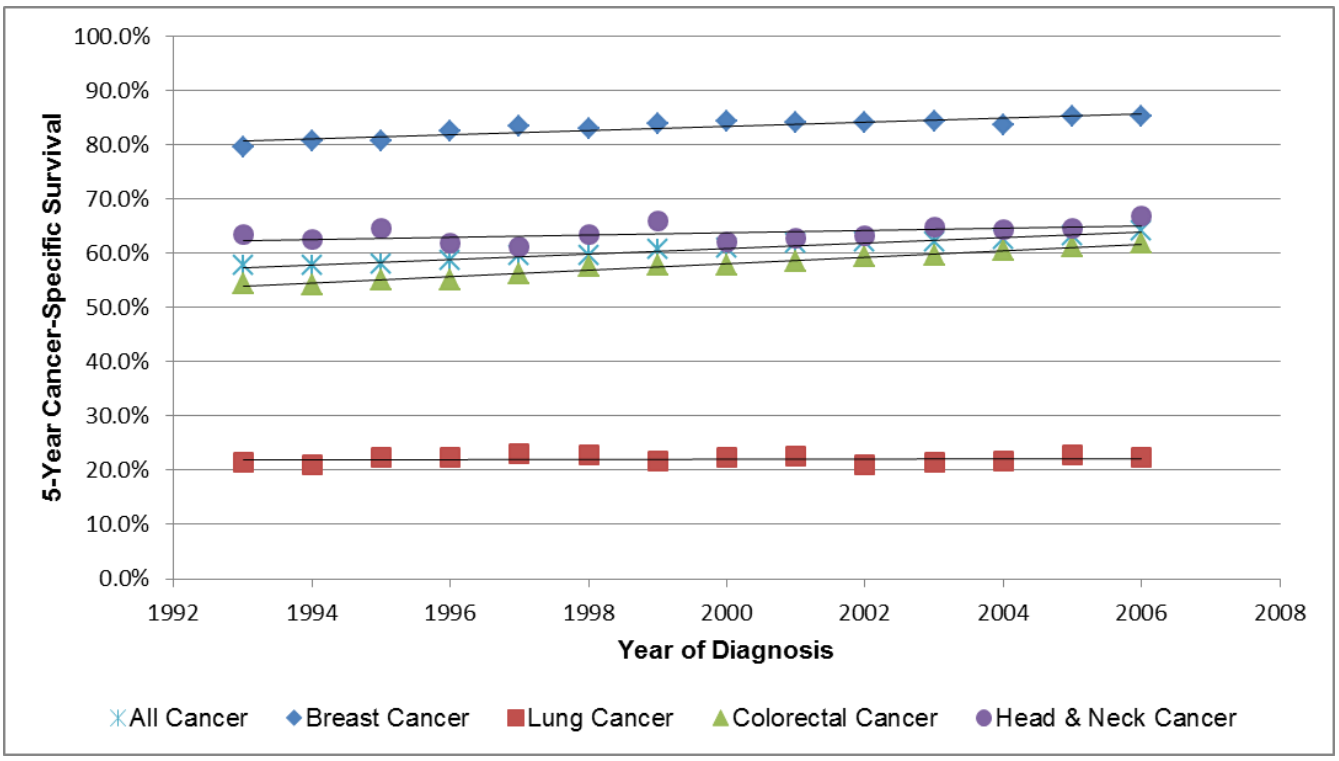

Figure 1. The 5-year cancer-specific survival, of all cancer cases in addition to breast, lung, head and neck, and colorectal cancer cases in Ontario from 1993 to 2006. The trend line was included for each cancer group.

3.3. Survival by SES

Figure 2 shows the SES trend in cancer survival among all cancer cases as well as the four cancer sites. Among all cancer cases, the 5-year cancer-specific survival was 57.3% (95% CI: 57.1 – 57.5) in the poorest SES quintile and 68.9% (95% CI: 68.6 – 69.1) in the richest SES quintile, an 11.6% difference. For breast cancer cases, the 5-year cancer specific survival was 81.6% (95% CI: 81.1 – 82.0) in the poorest SES quintile and 86.8% (95% CI: 86.3 – 87.2) in the richest SES quintile, a 5.2% difference. A similar, though less prominent, gradient was observed for lung cancer. The 5-year cancer-specific survival was 22.1% (95%

CI: 21.7 – 22.5) in the poorest SES quintile and 24.7% (95% CI: 23.9 – 25.5) in the richest SES quintile, a 2.6% difference. The 5-year cancer-specific survival among head and neck cancer cases showed the most noticeable SES trend relative to the other cancer sites. Survival was 60.4% (95% CI: 59.3 – 61.5) in the poorest SES quintile and 71.7% (95% CI: 70.3 – 73.2) in the richest SES quintile, an 11.3% difference. Finally, the 5-year cancer-specific survival of colorectal cancer was 57.1% (95% CI: 56.5 – 57.7) in the poorest SES quintile and 63.1% (95% CI: 62.3 – 63.8) in the richest SES quintile, a 6.0% difference.

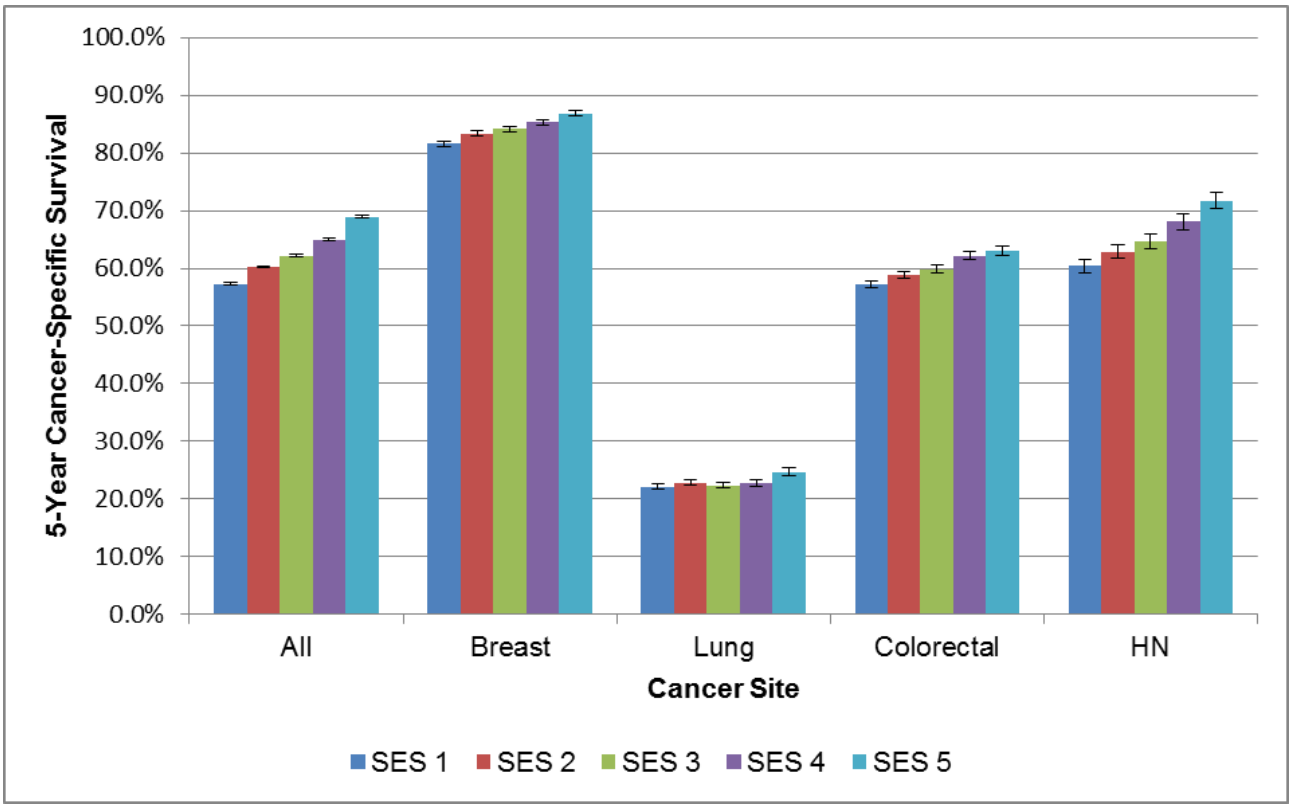

Figure 2. The 5-year cancer-specific survival, including 95% confidence intervals, of all cancer cases, as well as deaths due to breast, lung, head and neck, and colorectal cancer cases in Ontario from 1993 to 2006, separated by SES quintiles.

3.4. SES Effect by Year of Diagnosis

Improvement in cancer survival in Ontario between 1993 and 2006 was found to be dependent on the person's SES, as shown in Figure 3. The 5-year cancer-specific survival for patients residing in communities with a median household income within the poorest SES quintile (1^{st}) among all cancer cases improved by 3.5% from 55.3% (95% CI: 54.4 – 56.2) in 1993 to 58.8% (95% CI: 58.0 – 59.6) in 2006. For cancer cases residing in communities within the richest SES quintile (5^{th}), the 5-year cancer-specific survival improved by 8.6% from 63.4% (95% CI: 62.1 – 64.6) in 1993 to 72.0% (95% CI: 71.1 – 72.9) in 2006. Comparing the cancer-specific survival rates in the richest and poorest SES quintiles, the difference had widened from 8.1% in 1993 to 13.2% in 2006. Furthermore, cancer-specific survival among patients residing in communities with a median household income within the 2^{nd} – 4^{th} SES quintiles improved by 6.1%, 6.0%, and 7.7%, respectively. After controlling for age and sex in the Fine-Gray model, the hazard decreased by 3.1% (95% CI: 0.967 – 0.971) per year of diagnosis among patients in the richest SES quintile. In contrast, the hazard decreased by 1.2% (95% CI: 0.987 – 0.990) per year of diagnosis among patients in the poorest SES quintile (Table 2).

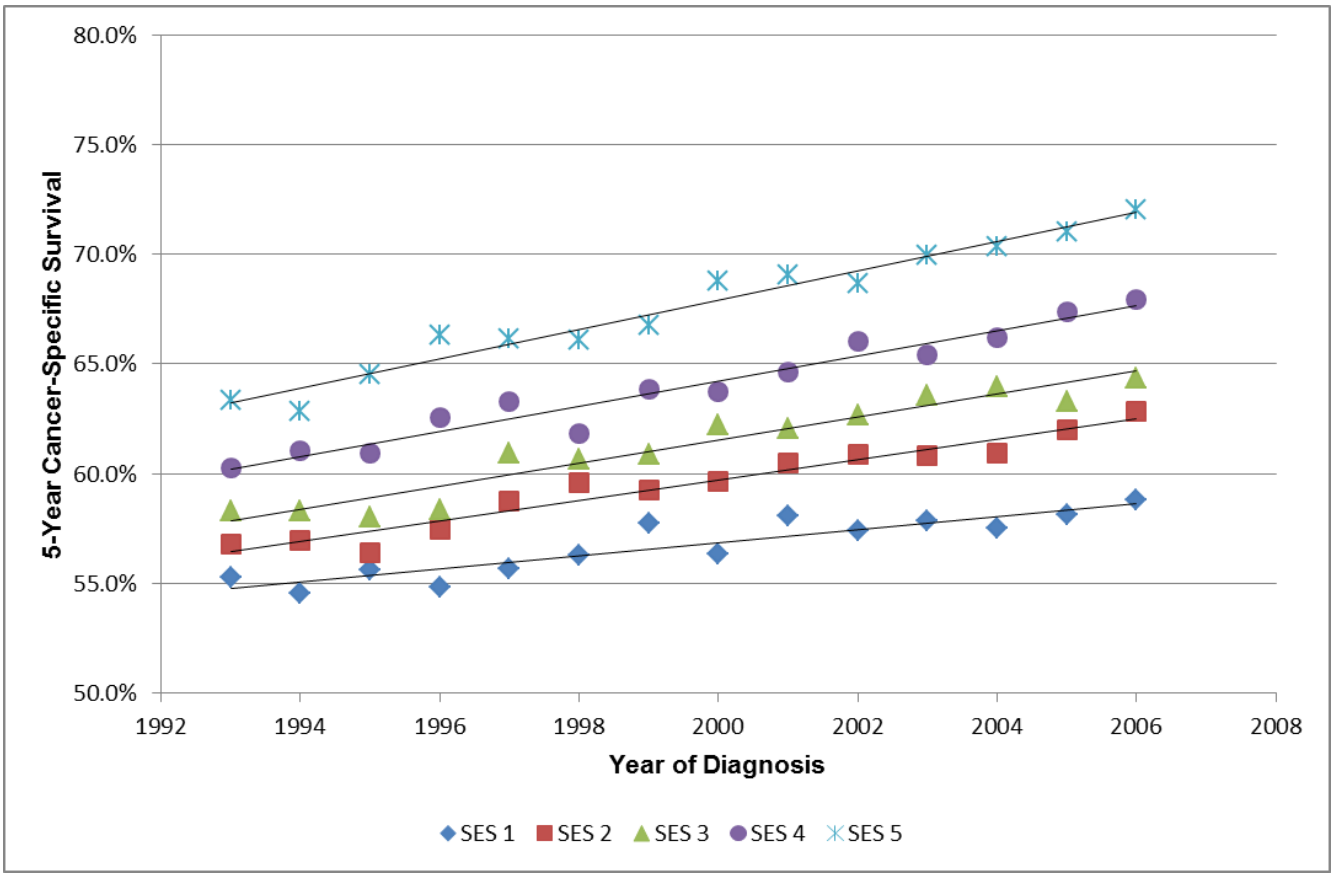

Figure 3. The 5-year cancer-specific survival of all cancer cases in Ontario from 1993 to 2006, separated by SES and year of diagnosis.

A similar temporal trend was found for specific cancer sites, but at different magnitudes (Figure 4). The 5-year cancer-specific survival for the poorest quintile among breast cancer cases improved by 2.2% from 80.4% (95% CI: 78.3 – 82.4) in 1993 to 82.6% (95% CI: 80.8 – 84.3) in 2006. In contrast, the 5-year cancer-specific survival among breast cancer cases in the richest quintile improved by 5.4% from 83.1% (95% CI: 80.7 – 85.4) in 1993 to 88.5% (95% CI: 86.8 – 90.0) in 2006 (Figure 4a). This resulted in a small, yet significant, increase from a 2.8% difference in survival between the poorest and richest quintiles in 1993 to a 5.9% difference in survival in 2006. After controlling

for age in the Fine-Gray model, the hazard decreased by 4.3% (95% CI: 0.951 – 0.964) per year of diagnosis among breast cancer patients in the richest SES quintile. In contrast, the hazard decreased by 2.0% (95% CI: 0.975 – 0.986) per year of diagnosis among breast cancer patients in the poorest SES quintile (Table 2).

For lung cancer patients in the poorest quintile, the 5-year cancer-specific survival changed by only 0.3% from 21.7% (95% CI: 20.0 – 23.6) in 1993 to 21.4% (95% CI: 19.7 – 23.1) in 2006. In contrast, for lung cancer cases in the richest quintile, the 5-year cancer-specific survival changed by 3.3% from 22.0% (95% CI: 19.0 – 25.4) in 1993 to 25.3% (95% CI: 22.6 – 28.3) in 2006 (Figure 4b). This resulted in a significant increase in cancer survival between the poorest and richest quintiles from a 0.3% difference in 1993 to a 3.9% difference in 2006. Lung cancer patients in the 2^{nd} – 4^{th} SES quintiles showed no substantial change in survival during the study period. After controlling for age and sex in the Fine-Gray model, the hazard decreased by 1.4% (95% CI: 0.982 – 0.990) per year of diagnosis among lung cancer patients in the richest SES quintile. In contrast, the hazard decreased by 0.3% (95% CI: 0.995 – 1.000) per year of diagnosis among lung cancer patients in the poorest SES quintile (Table 2).

For colorectal cancer patients, the 5-year cancer-specific survival among those in the poorest quintile increased by 3.4% from 56.2% (95% CI: 53.8 – 58.6) in 1993 to 59.6% (95% CI: 57.3 – 61.8) in 2006. In contrast, for those in richest quintile, the 5-year cancer-specific survival increased by 12.3% from 56.5% (95% CI: 52.9 – 60.2) in 1993 to 68.8% (95% CI: 66.2 – 71.4) in 2006 (Figure 4c). This resulted in a significant increase in cancer survival between the poorest and richest quintiles from a 0.3% difference in 1993 to a 9.2% difference in 2006. After controlling for age and sex in the Fine-Gray

model, the hazard decreased by 3.7% (95% CI: 0.958 – 0.968) per year of diagnosis among colorectal cancer patients in the richest SES quintile. In contrast, the hazard decreased by 1.8% (95% CI: 0.978 – 0.985) per year of diagnosis among colorectal cancer patients in the poorest SES quintile (Table 2). For head and neck cancer, the trend was unclear due to the relatively small sample sizes resulting in large uncertainty (Figure 4d).

The effects of SES, year of diagnosis, and their interaction effect on the time to cancer-specific death using the Fine-Gray model, controlling for age and sex, are reported in Table 3. It includes results from the models based on all cancer cases and based on breast, lung, head and neck, and colorectal cancer cases, separately. Sex is not controlled in the model for breast cancer cases because of few male breast cancer cases in the data. Year of diagnosis was treated as a continuous variable in the model. For all cancer cases as well as breast, lung, head and neck, and colorectal cancer cases, the hazard of cancer-specific death decreases significantly over the study period in all SES groups (Table 2). The rate of the decrease was highest in the richest group and lowest in the poorest group, and the Fine-Gray model shows that the differences in the rate among different SES groups are significant.

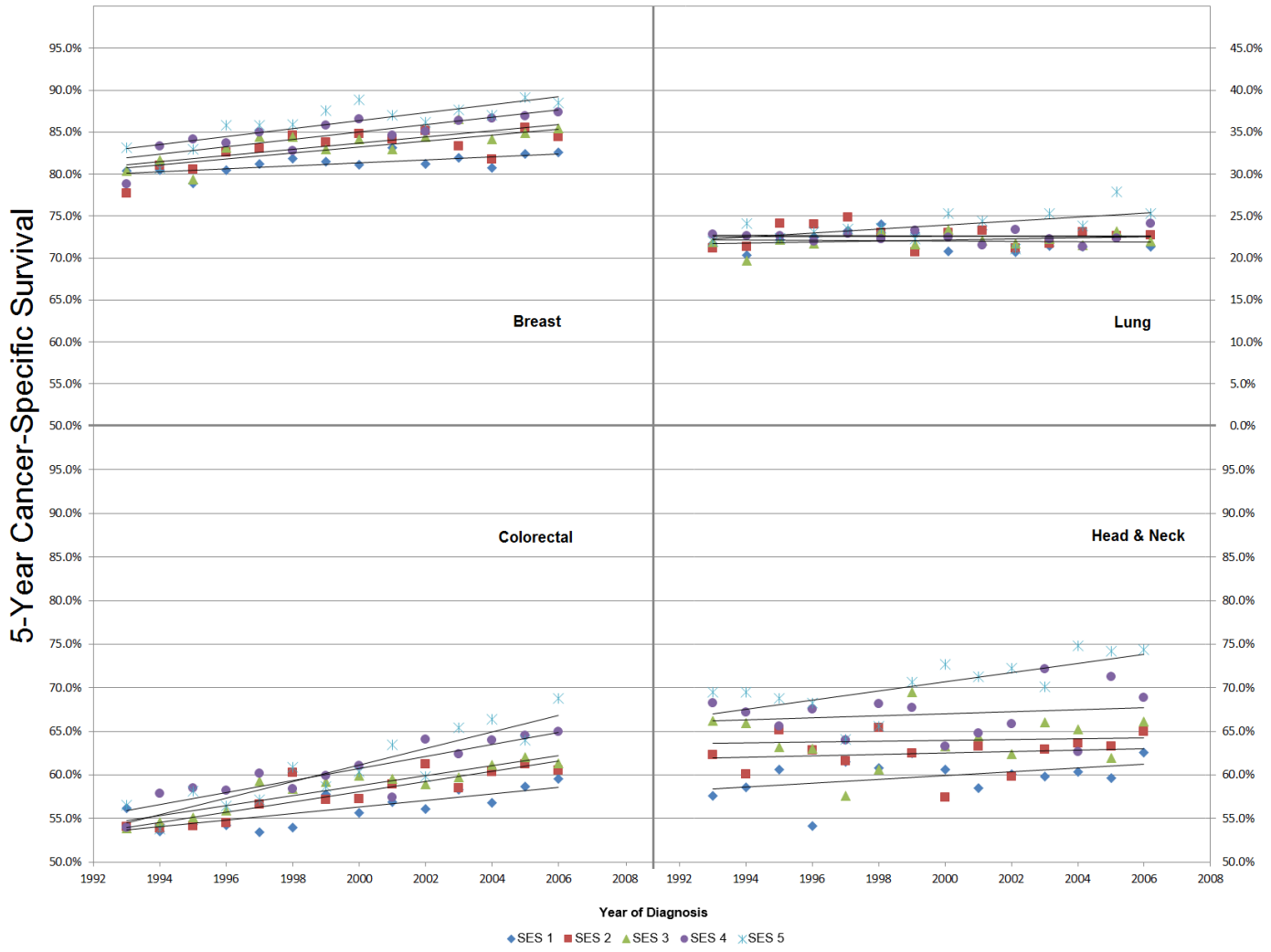

Figure 4. The 5-year cancer-specific survival of breast cancer (a), lung cancer (b), colorectal cancer (c), and head & neck cancer (d) in Ontario from 1993 to 2006, separated by SES and year of diagnosis.

Table 2. The hazard ratio per year of diagnosis, separated by SES, for cancer cases in Ontario diagnosed between 1993 and 2009. The hazard ratios and their 95% confidence intervals were calculated from the Fine-Gray model

Site	SES	HR	95% Confidence Intervals	
All				
	1	0.988	0.987	0.990
	2	0.983	0.982	0.985
	3	0.980	0.978	0.981
	4	0.977	0.975	0.979
	5	0.969	0.967	0.971
Breast				
	1	0.980	0.975	0.986
	2	0.974	0.968	0.979
	3	0.968	0.962	0.973
	4	0.966	0.960	0.972
	5	0.957	0.951	0.964
Lung				
	1	0.997	0.995	1.000
	2	0.997	0.995	1.000
	3	0.995	0.992	0.997
	4	0.996	0.992	0.999
	5	0.986	0.982	0.990
Colorectal				
	1	0.982	0.978	0.985
	2	0.978	0.975	0.982
	3	0.975	0.971	0.979
	4	0.971	0.966	0.975
	5	0.963	0.958	0.968
Head & Neck				
	1	0.990	0.983	0.996
	2	0.988	0.981	0.995
	3	0.985	0.978	0.993
	4	0.979	0.970	0.988
	5	0.969	0.958	0.979

Table 3. Summary of Fine-Gray model for all cancer cases in addition to breast, lung, head and neck, and colorectal cancer cases in Ontario

Site	Parameter	Group	Estimate	p-value	Hazard Ratio
All					
	SES	1	0.10975	*<.0001*	
		2	0.07675	*<.0001*	
		3	0.06418	*<.0001*	
		4	0.03002	*0.0085*	
	Year of Diagnosis		-0.03145	*<.0001*	
	Year of Diagnosis x SES	1	0.01968	*<.0001*	
	Year of Diagnosis x SES	2	0.01457	*<.0001*	
	Year of Diagnosis x SES	3	0.01098	*<.0001*	
	Year of Diagnosis x SES	4	0.0082	*<.0001*	
	SEX	Female	-0.04869	*<.0001*	0.952
	AGE	<50	-1.09359	*<.0001*	0.335
		50-59	-0.65857	*<.0001*	0.518
		60-69	-0.46141	*<.0001*	0.630
		70-79	-0.23688	*<.0001*	0.789
Breast					
	SES	1	0.0529	0.1303	
		2	0.03253	0.3682	
		3	0.04568	0.2118	
		4	-0.00303	0.9363	
	Year of Diagnosis		-0.04368	*<.0001*	
	Year of Diagnosis x SES	1	0.0241	*<.0001*	
	Year of Diagnosis x SES	2	0.01703	*0.0002*	
	Year of Diagnosis x SES	3	0.01103	*0.0158*	
	Year of Diagnosis x SES	4	0.00929	0.0508	
	AGE	<50	-0.64488	*<.0001*	0.525
		50-59	-0.69791	*<.0001*	0.498
		60-69	-0.58719	*<.0001*	0.556
		70-79	-0.33706	*<.0001*	0.714
Lung					
	SES	1	-0.00547	0.8002	
		2	-0.02957	0.1852	
		3	-0.00886	0.6979	
		4	-0.02888	0.2326	
	Year of Diagnosis		-0.0141	*<.0001*	

	Year of Diagnosis x SES	1	0.01128	*<.0001*	
	Year of Diagnosis x SES	2	0.01135	*<.0001*	
	Year of Diagnosis x SES	3	0.00871	*0.0001*	
	Year of Diagnosis x SES	4	0.00976	*0.0001*	
	SEX	Female	-0.11239	*<.0001*	0.894
	AGE	<50	-0.37455	*<.0001*	0.688
		50-59	-0.24243	*<.0001*	0.785
		60-69	-0.21057	*<.0001*	0.810
		70-79	-0.12318	*<.0001*	0.884
Head & Neck					
	SES	1	0.15557	*0.0088*	
		2	0.10262	0.0922	
		3	0.05735	0.3593	
		4	0.02029	0.7575	
	Year of Diagnosis		-0.03196	*<.0001*	
	Year of Diagnosis x SES	1	0.02158	*0.001*	
	Year of Diagnosis x SES	2	0.01965	*0.004*	
	Year of Diagnosis x SES	3	0.01725	*0.013*	
	Year of Diagnosis x SES	4	0.0108	0.138	
	SEX	Female	-0.11748	*<.0001*	0.889
	AGE	<50	-0.90383	*<.0001*	0.405
		50-59	-0.45215	*<.0001*	0.636
		60-69	-0.28218	*<.0001*	0.754
		70-79	-0.14003	*<.0001*	0.869
Colorectal					
	SES	1	-0.02266	0.416	
		2	-0.04186	0.1427	
		3	-0.03974	0.1743	
		4	-0.05874	0.0557	
	Year of Diagnosis		-0.03779	*<.0001*	
	Year of Diagnosis x SES	1	0.01916	*<.0001*	
	Year of Diagnosis x SES	2	0.016	*<.0001*	
	Year of Diagnosis x SES	3	0.01267	*<.0001*	
	Year of Diagnosis x SES	4	0.00812	*0.016*	
	SEX	Female	-0.06673	*<.0001*	0.935
	AGE	<50	-0.55141	*<.0001*	0.576
		50-59	-0.50594	*<.0001*	0.603
		60-69	-0.41254	*<.0001*	0.662
		70-79	-0.29245	*<.0001*	0.746

4. DISCUSSION

It was hypothesized that the difference in magnitude in cancer-specific survival between the richest and poorest SES groups between 1993 and 2009 would not change during the study period. This study demonstrates that while cancer-specific survival in Ontario has significantly improved between 1993 and 2006, the change in the hazard over time has been found to be dependent on the patient's SES. This was found across all cancer cases in addition to breast, lung, head and neck, and colorectal cancer cases.

For lung cancer cases, the interaction effect of SES and year of diagnosis was small. The effect is only observed between the richest SES group and the poorer SES groups. For colorectal and breast cancer cases, there is a great degree of separation between all the SES groups over time. This temporal trend is significant for all cancer cases as well as breast, lung, head and neck, and colorectal cancer cases after adjusting for age and sex.

The temporal trend in Ontario may be due to cancer patients of higher SES having differential access to treatment/screening and/or better education during the study period[1]. As well, a patient's SES may be associated with the incidence of comorbidities during treatment[10,11]. However, it is difficult to assess the precise causes of this temporal trend at this time.

The distribution of median household income in each DA/EA was not accounted for in this study. Since the highest SES quintile could become more skewed to the right over time as a result of the increasing income disparity in Ontario, the trend could simply be an artifact of this change. However, Figures 3-4 do not show sudden changes in cancer survival

between each census period (1997 – 1998, 2002 – 2003), as would be observed if income disparity was playing a major role.

The OCR did not include the lack of stage, ethnicity, comorbidity, and smoking status. Stage, in particular, was only available for more recent cases and, as a result, could not be included into the model to examine their potential causal association with the temporal trend. Cases with missing income data cannot be included into the model. The potential role of treatment access and adherence was not examined. As well, this study did not look into the possible effect of changes in screening procedures and medical technologies over the study period. Overall, this study did not examine potential causal relationships with regards to the temporal trend. As well, it should be noted that the quality of the cause of death information in the OCR is not perfect[17]. Although, it should also be noted that the error rate for cause of death information in the OCR has improved over time.

It should be noted that this study assumed that median household income in a neighborhood reflects the patient's individual income and lifestyle. This study does not suggest that the median household income of the community equates to the patient's individual income or education. However, using household income at the individual level would result in a violation of the patient's privacy. This study reflects changes in cancer survival among Ontario cases only.

Future research should look into the effect of stage as well as the patient's smoking status on the trend. Even though stage has not been found to significantly affect the SES trend in Ontario in prior studies, its effect on the trend over time may still exist. As well, due to its availability in more recent cases, it is certainly possible to include stage for future

studies assuming the data from the 2011 census is reliable. Finally, the effect of the distribution of median household income in each DA/EA on the trend should be examined.

1. SUMMARY

The purpose of this practicum was to examine the change in the SES effect on cancer survival in Ontario from 1993 to 2009. It was found that the difference in cancer survival between patients in the richest SES group and patients in the poorer SES groups has increased during that time span. The change in the hazard during the study period was found to be significantly associated with the patient's SES. It is uncertain whether this change is due to factors such as the increasing income disparity, differential access to treatment, screening procedures, and/or other causes.

2. REFERENCES

1. Cella DF, Orav EJ, Kornblith AB, Holland JC, Silberfarb PM, Lee KW, Comis RL, Perry M, Cooper R, Maurer LH, et al. (1991). Socioeconomic status and cancer survival. *J Clin Oncol*. 9(8): 1500-9.
2. Mackillop WJ, Zhang-Salomons J, Groome PA, Paszat L, Holowaty E. (1997). Socioeconomic status and cancer survival in Ontario. *J Clin Oncol*. 15(4): 1680-9.
3. Yu XQ. (2009). Socioeconomic disparities in breast cancer survival: relation to stage at diagnosis, treatment and race. *BMC Cancer*. 9: 364.
4. Byers TE, Wolf HJ, Bauer KR, Bolick-Aldrich S, Chen VW, Finch JL, Fulton JP, Schymura MJ, Shen T, Van Heest S, Yin X; Patterns of Care Study Group. (2008).

The impact of socioeconomic status on survival after cancer in the United States : findings from the National Program of Cancer Registries Patterns of Care Study. *Cancer*. 113(3): 582-91.

5. McDonald JT, Johnson-Obaseki S, Hwang E, Connell C, Corsten M. (2014). The relationship between survival and socio-economic status for head and neck cancer in Canada. *J Otolaryngol Head Neck Surg*. 43: 2.
6. Ramsay SE, Morris RW, Whincup PH, Papacosta AO, Lennon LT, Wannamethee SG. (2014). Time trends in socioeconomic inequalities in cancer mortality: results from a 35 year prospective study in British men. *BMC Cancer*. 14: 474.
7. Niu X, Roche LM, Pawlish KS, Henry KA. (2013). Cancer survival disparities by health insurance status. *Cancer Med*. 2(3): 403-11.
8. Fine JP, Gray RJ. (1999). A Proportional Hazards Model for the Subdistribution of a Competing Risk. *J Am Stat Assoc*. 94: 496-509
9. Geskus RB. (2011). Cause-specific cumulative incidence estimation and the fine and gray model under both left truncation and right censoring. *Biometrics*. 67(1): 39-49.
10. Schrijvers CT, Coebergh JW, Mackenbach JP. (1997). Socioeconomic status and comorbidity among newly diagnosed cancer patients. *Cancer*. 80(8): 1482-8.
11. Grose D, Morrison DS, Devereux G, Jones R, Sharma D, Selby C, Docherty K, McIntosh D, Louden G, Nicolson M, McMillan DC, Milroy R; Scottish Lung Cancer Forum. (2014). Comorbidities in lung cancer: prevalence, severity and links with socioeconomic status and treatment. *Postgrad Med J*. 90(1064): 305-10.

12. Booth CM, Li G, Zhang-Salomons J, Mackillop WJ. (2010). The impact of socioeconomic status on stage of cancer at diagnosis and survival: a population-based study in Ontario, Canada. *Cancer*. 116(17): 4160-7.
13. Boyd C, Zhang-Salomons JY, Groome PA, Mackillop WJ. (1999). Associations between community income and cancer survival in Ontario, Canada, and the United States. *J Clin Oncol*. 17(7): 2244-55.
14. Schrijvers CT, Mackenbach JP. (1994). Cancer patient survival by socioeconomic status in seven countries: a review for six common cancer sites [corrected]. *J Epidemiol Community Health*. 48(5): 441-6.
15. Aizer AA, Chen MH, McCarthy EP, Mendu ML, Koo S, Wilhite TJ, Graham PL, Choueiri TK, Hoffman KE, Martin NE, Hu JC, Nguyen PL. (2013). Marital status and survival in patients with cancer. *J Clin Oncol*. 31(31): 3869-76.
16. Kaplan EL, Meier P. (1958). Nonparametric Estimation from Incomplete Observations. *J Amer Statist Assoc*. 53(282): 457-481.
17. Hall S, Schulze K, Groome P, Mackillop W, Holowaty E. (2006). Using cancer registry data for survival studies: the example of the Ontario Cancer Registry. *J Clin Epidemiol*. 59(1): 67-76.
18. Schoenfeld D. (1982). Partial residuals for the proportional hazards regression model. *Biometrika*. 69(1): 239-241.

3. APPENDIX A – OVERALL SURVIVAL BY YEAR OF DIAGNOSIS

From 1993 to 2009, 3-year overall survival for all cancer cases has significantly improved from 55.0% (95% CI: 54.5 – 55.4) to 63.8% (95% CI: 63.4 – 64.2). A similar trend was seen in all 4 major cancer sites: the 3-year survival improved from 81.1% (95% CI: 80.1 – 82.0) in 1993 to 87.1% (95% CI: 86.4 – 87.8) in 2009 for breast cancer; from 17.8% (95% CI: 16.8 – 18.7) to 21.8% (95% CI: 20.9 – 22.7) for lung cancer; from 60.1% (95% CI: 57.7 – 62.5) to 67.9% (95% CI: 65.8 – 69.8) for head and neck cancer; and from 52.9% (95% CI: 51.6 – 54.2) to 65.3% (95% CI: 64.3 – 66.3) for colorectal cancer (Figure 8).

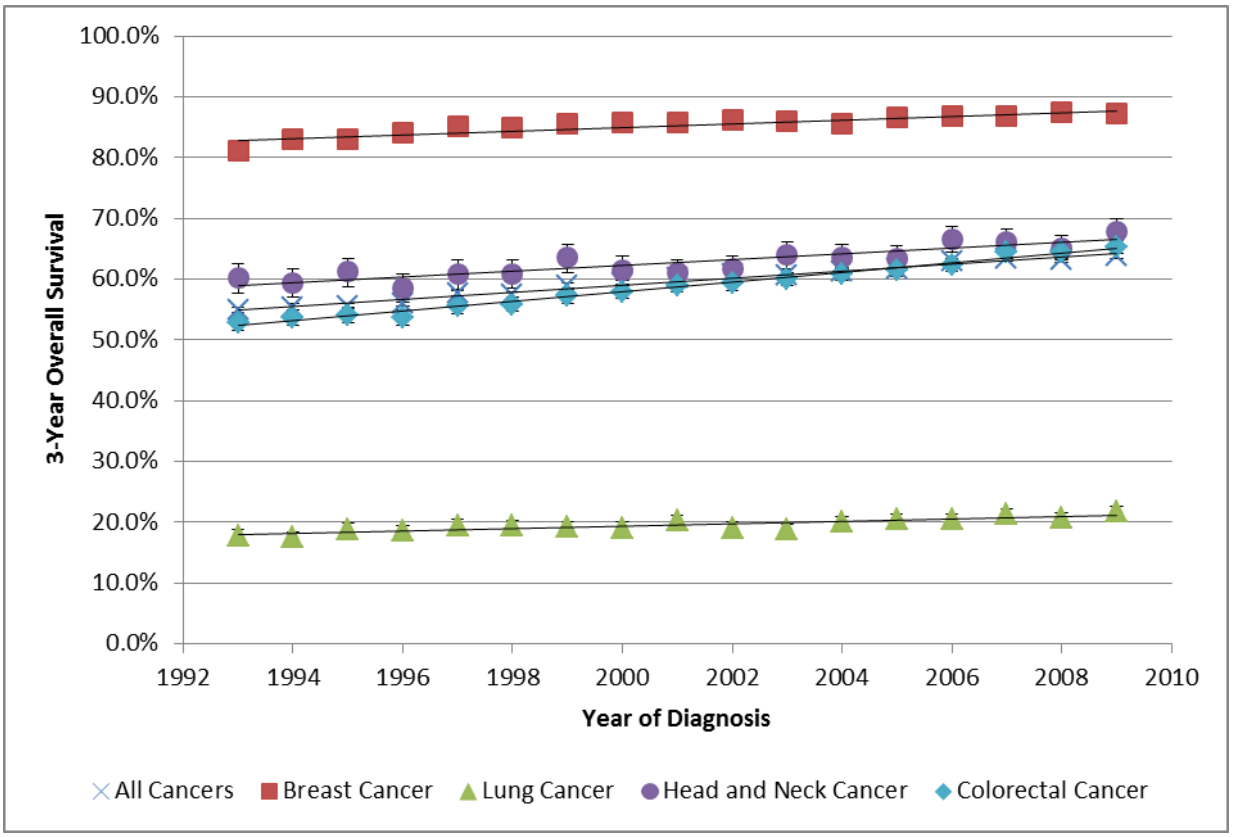

Figure 5. The 3-year overall survival, of all cancer cases in addition to breast, lung, head and neck, and colorectal cancer cases in Ontario from 1993 to 2008. These values were calculated from the Kaplan-Meier estimator. The trend line was included for each cancer group.

4. APPENDIX B – SES TREND AND OVERALL SURVIVAL

Figure 9 shows the difference in the SES trend between overall survival and cancer-specific survival. During the study period, the 3-year overall survival of the cancer patients in the SES groups were, from the richest to the poorest: 69.1% (95% CI: 68.8 – 69.3), 64.2% (95% CI: 64.0 – 64.4), 60.8% (95% CI: 60.6 – 61.1), 58.3% (95% CI: 58.1 – 58.5), and 54.3% (95% CI: 54.1 – 54.6).

By comparison, the 3-year cancer-specific survival of the cancer patients in the SES groups were, from the richest to the poorest: 73.3% (95% CI: 73.0 – 73.5), 69.5% (95% CI: 69.2 – 69.7), 66.8% (95% CI: 66.6 – 67.0), 65.0% (95% CI: 64.8 – 65.2), and 62.1% (95% CI: 61.8 – 62.3). Overall, there was a slightly more noticeable SES trend in 3-year overall survival between the richest and poorest SES groups (14.7% difference) than in 3-year overall survival (11.2% difference) (Figure 9).

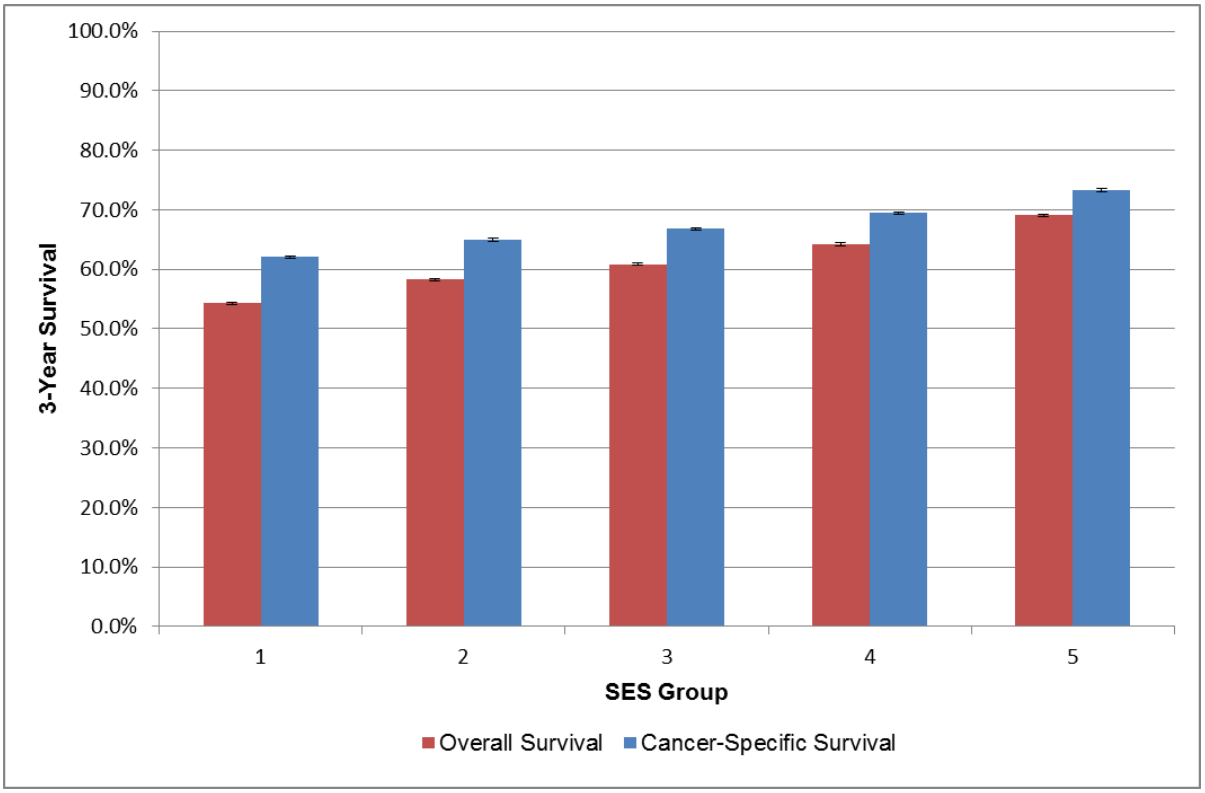

Figure 6. The 3-year overall survival and cancer-specific survival, including their 95% confidence intervals, for all cancer cases in Ontario from 1993 to 2008, separated by SES quintiles.

5. APPENDIX C – FINE-GRAY REGRESSION

Fine and Gray offered a competing risks regression method that models covariates directly to the cumulative incidence function[8]. For any type of event (*k*), this method focuses on the hazard associated with the cumulative incidence function, $I_k(t; x)$, which is the probability that an event of type *k* (cancer vs. non-cancer) has occurred by time *t* (months). They did this through the subdistribution hazard function, $h^*_k(t)$:

$$h^*_k(t; x) = h^*_{k0}(t)\, e^{\beta x}$$

This resembles the standard Cox model:

$$h(t; x) = h_0(t)\, e^{\beta x}$$

The cumulative incidence of the subdistribution is the function such that:

$$I_k(t; x) = 1 - \exp\{-\int_0^t h *_k (u, x) du\}$$

This shows that the cumulative incidence function depends on the hazard for cause *k*. As such, the subdistribution hazard can also be defined as follows:

$$h *_k (t; x) = -\frac{d}{dt}\left(1 - I_k(t; x)\right)$$

This directly relates the covariate effects to the cumulative incidence function. Fine and Gray imposed a proportional hazards assumption on the subdistribution hazards. Estimation of the covariate coefficients in the Fine-Gray model follows the partial likelihood approach, as used in the Cox model. As well, the subdistribution hazard differs from the cause-specific hazard, $h_k(t)$, the instantaneous rate that an event *k* has occurred by time *t*, by how each handles competing events. In the cause-specific hazard, the risk

set decreases at each time point whenever a competing event occurs. In the subdistribution hazard, a person who has experienced the competing event remains in the risk set.

Schoenfeld residuals, when plotted with time, can assess whether the proportionality hazards assumption is met in the Fine-Gray model. The Schoenfeld residual is defined as the covariate value for the patient that failed minus the expected value. The expected value at time *t* is the weighted average of the covariate, weighted by the probability of failure for each patient in the risk set at time t[18]. If a linear correlation between the Schoenfeld residuals and time is present, then the covariate is likely time-dependent and it violates the proportional hazards assumption.

6. APPENDIX D – MISSING DATA

Since either participation rates for each census is never 100% or dissemination areas exist with populations smaller than the required cell size for reporting, missing data is unavoidable. In addition, some cancer cases do no report a postal code, making any linkage difficult. As a result, there is usually a 3-6% rate of cases without a reported median household income in each year of diagnosis (Figure 7) and a 4.0% missing percentage during the study period (Table 1).

As Figure 7 shows, cases linked to the 2001 census (1998 – 2002) showed a higher missing rate than cases linked to the 1996 (1993 – 1997) and 2006 censuses (2003 – 2009). This discrepancy may be due to slightly greater difficulties linking the 2001 census to the OCR.

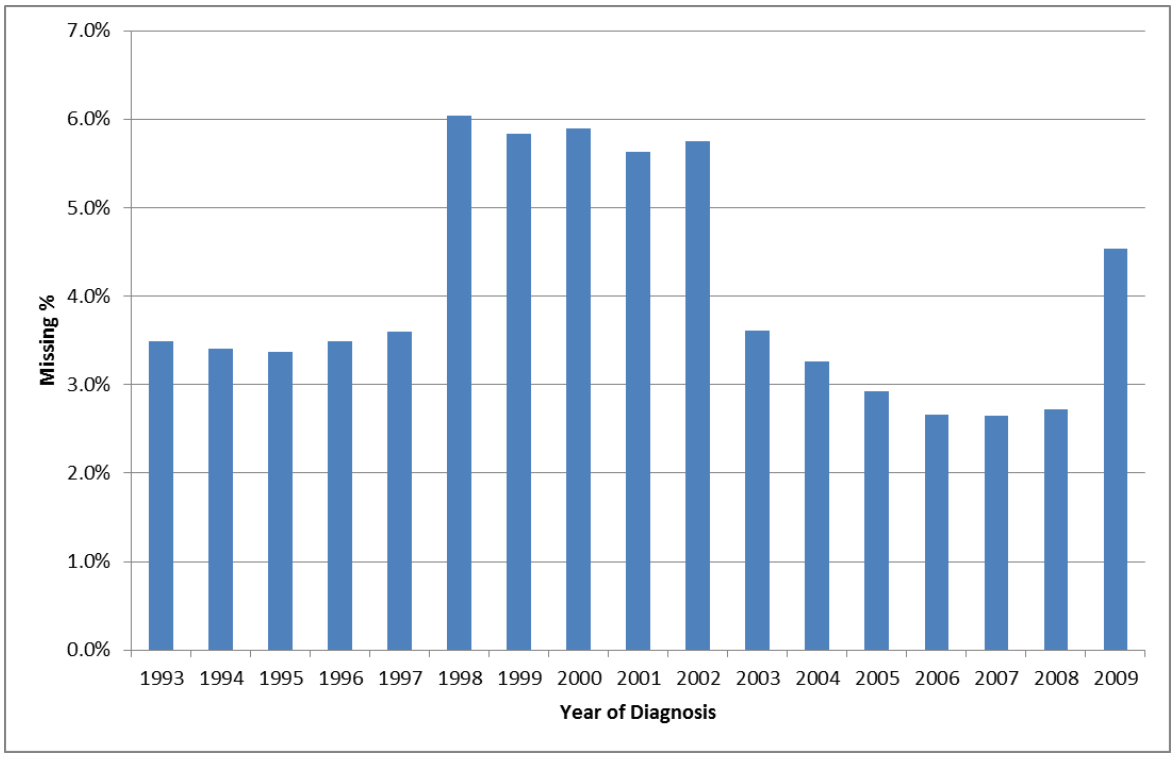

Figure 7. The percentage of missing cases in each year of diagnosis in Ontario from 1993 to 2009.

As for the effect of missing data on survival, cases without a reported median household income were compared to cases with median household income. What was found was that the 3-year survival of cases without income data was always lower than for cases with income data (Figure 8). The reason for this may be due to cases dying before giving out much of their personal information, including their postal code.

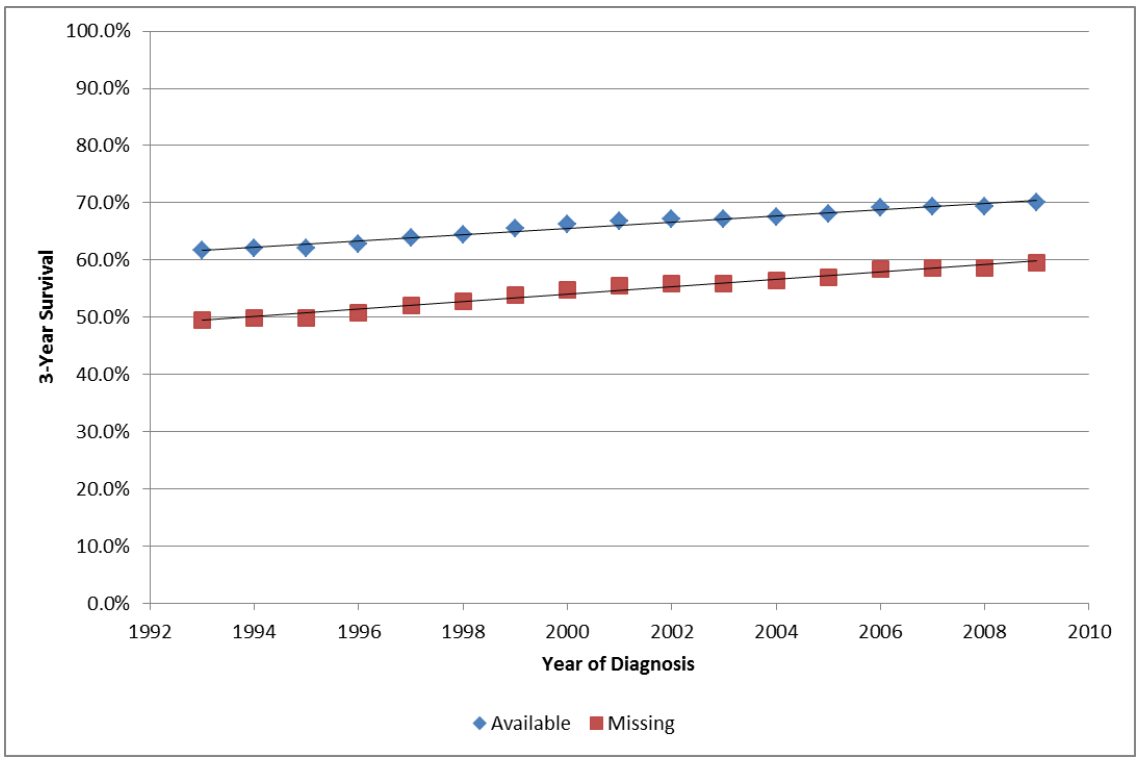

Figure 8. The 3-year overall survival, including 95% confidence intervals, for all cancer cases in Ontario with a reported median household income data (Available) and cases without a reported median household income (Missing), from 1993 – 2009. Calculated using the Kaplan-Meier estimator.

7. APPENDIX E – SCHOENFELD RESIDUAL PLOTS

To assess whether proportionality has been met, the Fine-Gray model outputted Schoenfeld residuals for each covariate including the interactions. None of the plots showed any interaction between the residuals and time, thereby meeting the proportional hazards assumption (Figures 9-22). Due to the large sample size of the dataset, the SAS procedure could not produce the LOESS curves for the plots.

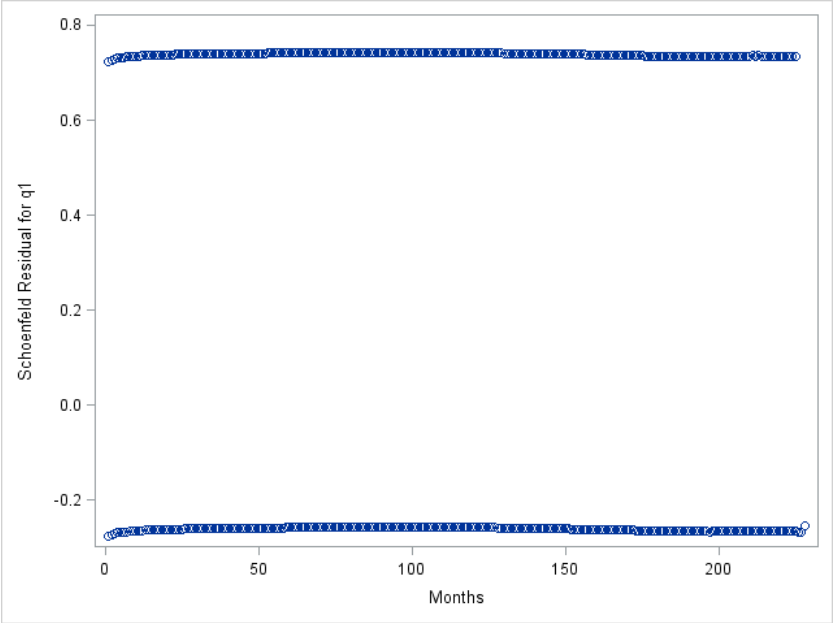

Figure 9. Graphic analysis of proportional hazards assumption using Schoenfeld residuals (SES group 1)

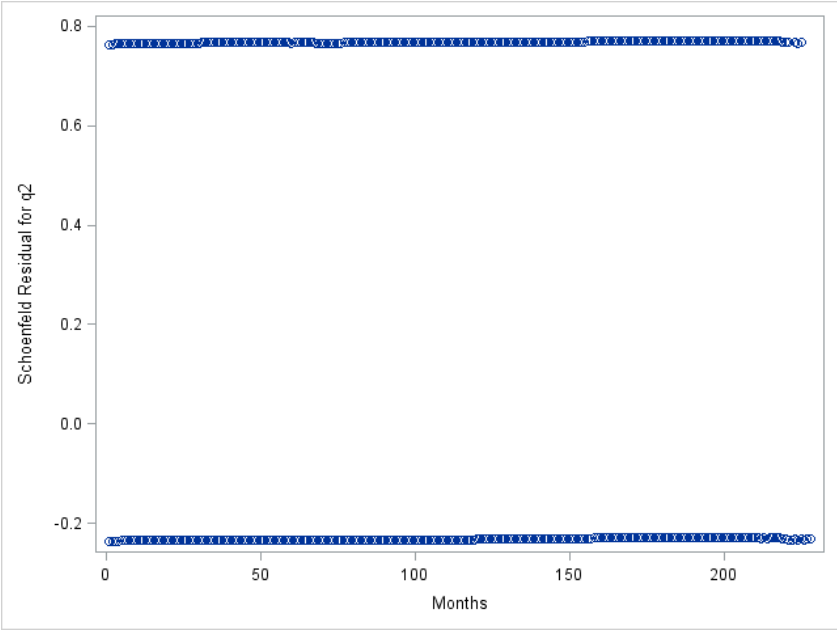

Figure 10. Graphic analysis of proportional hazards assumption using Schoenfeld residuals (SES group 2)

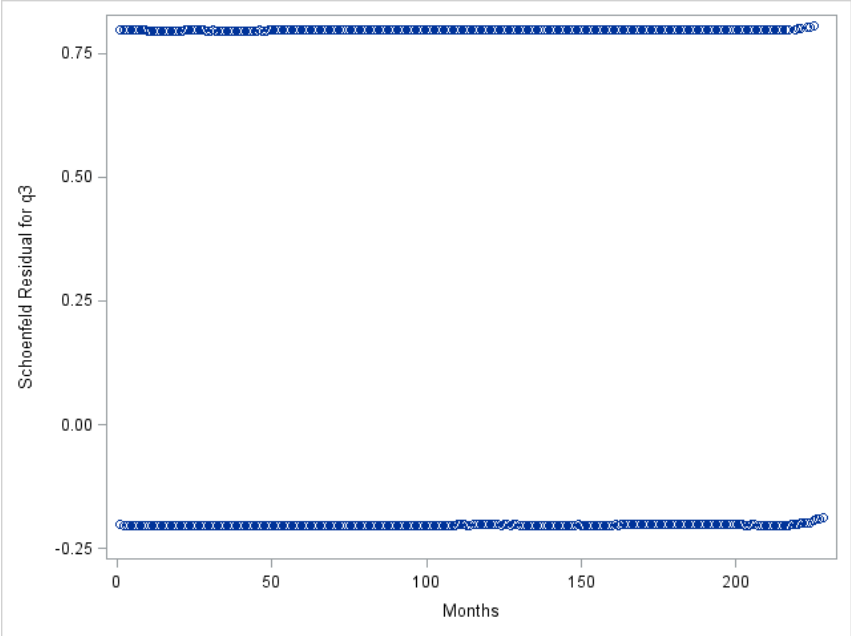

Figure 11. Graphic analysis of proportional hazards assumption using Schoenfeld residuals (SES group 3)

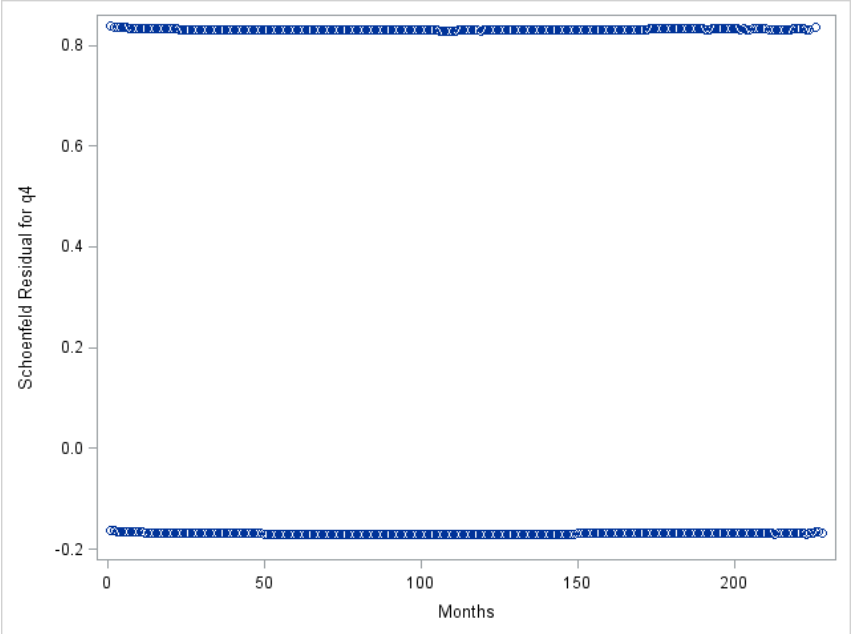

Figure 12. Graphic analysis of proportional hazards assumption using Schoenfeld residuals (SES group 4)

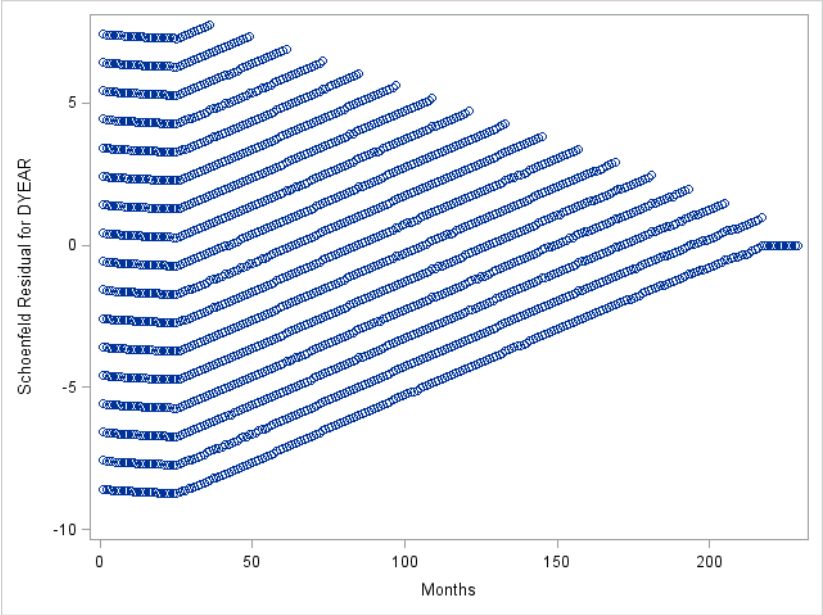

Figure 13. Graphic analysis of proportional hazards assumption using Schoenfeld residuals (Year of diagnosis)

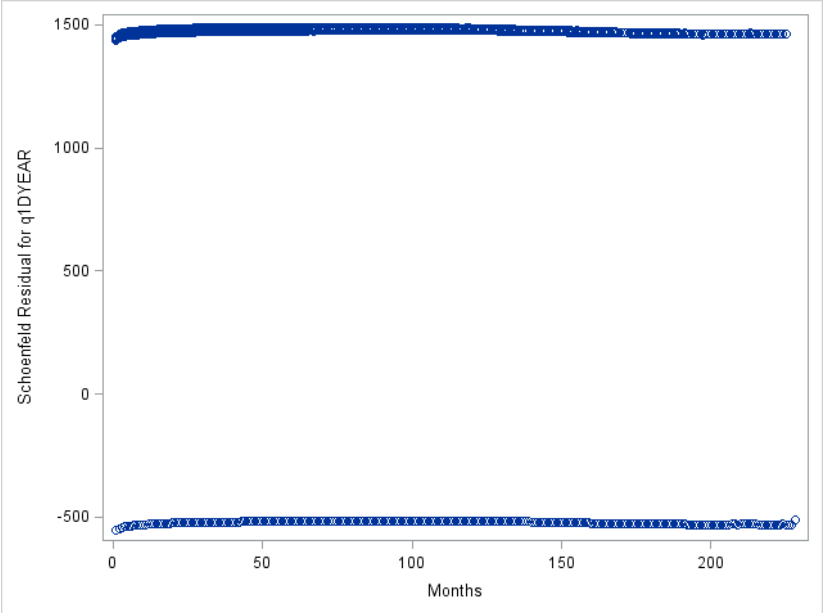

Figure 14. Graphic analysis of proportional hazards assumption using Schoenfeld residuals (SES group 1 x Year of diagnosis)

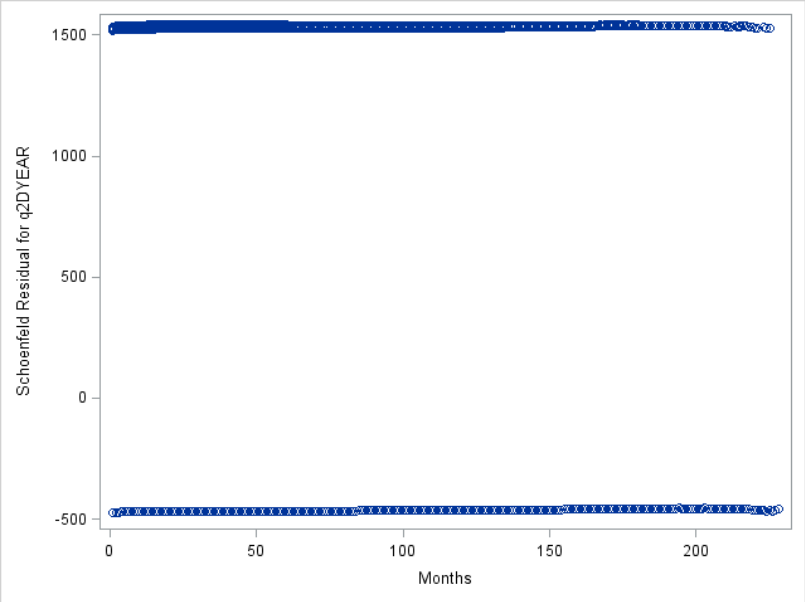

Figure 15. Graphic analysis of proportional hazards assumption using Schoenfeld residuals (SES group 2 x Year of diagnosis)

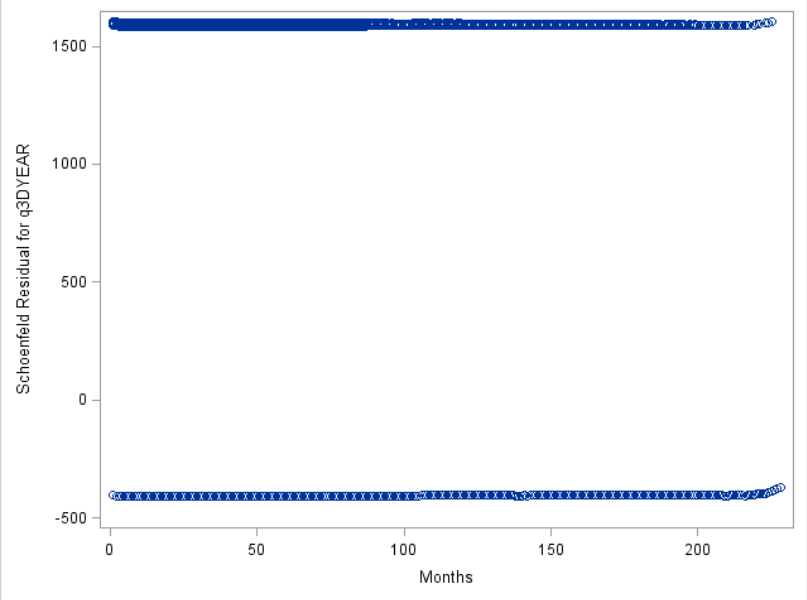

Figure 16. Graphic analysis of proportional hazards assumption using Schoenfeld residuals (SES group 3 x Year of diagnosis)

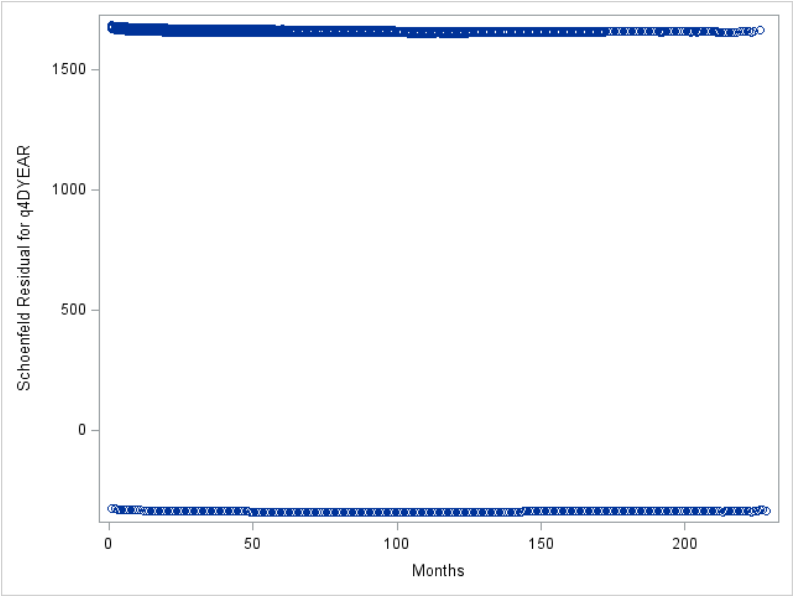

Figure 17. Graphic analysis of proportional hazards assumption using Schoenfeld residuals (SES group 4 x Year of diagnosis)

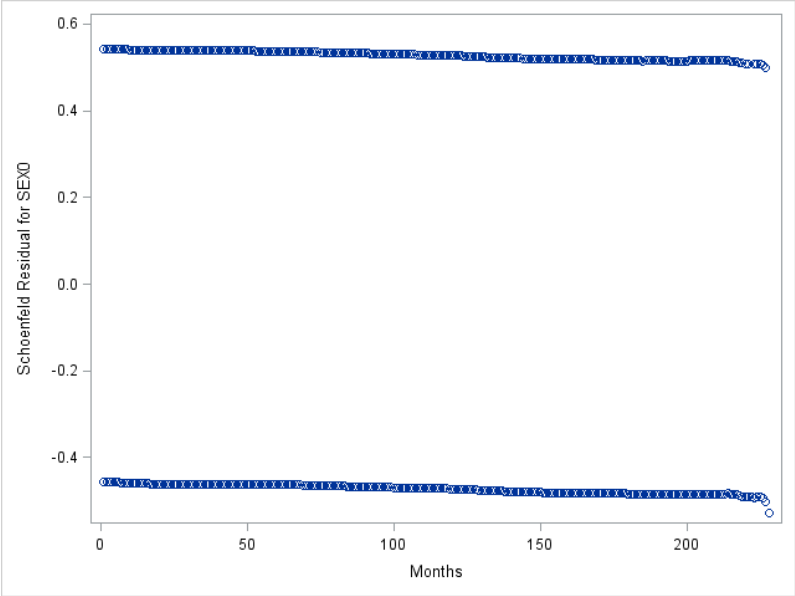

Figure 18. Graphic analysis of proportional hazards assumption using Schoenfeld residuals (Sex)

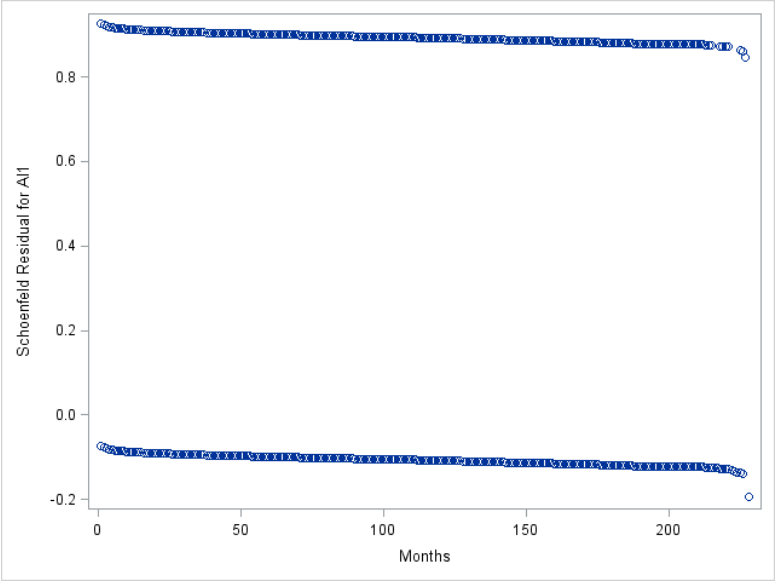

Figure 19. Graphic analysis of proportional hazards assumption using Schoenfeld residuals (Age group 1)

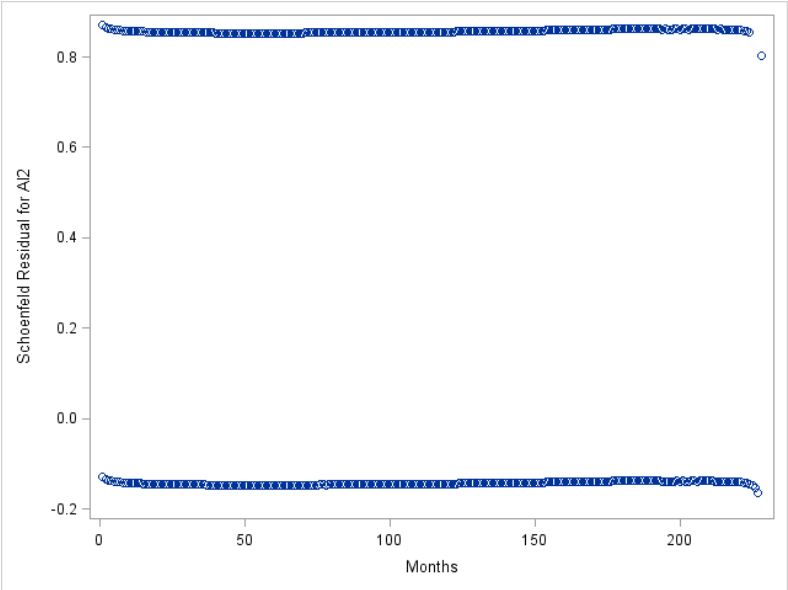

Figure 20. Graphic analysis of proportional hazards assumption using Schoenfeld residuals (Age group 2)

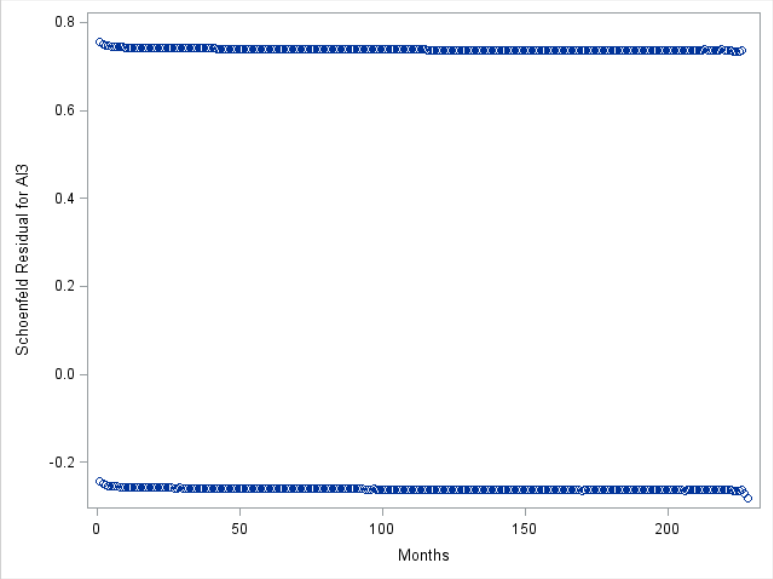

Figure 21. Graphic analysis of proportional hazards assumption using Schoenfeld residuals (Age group 3)

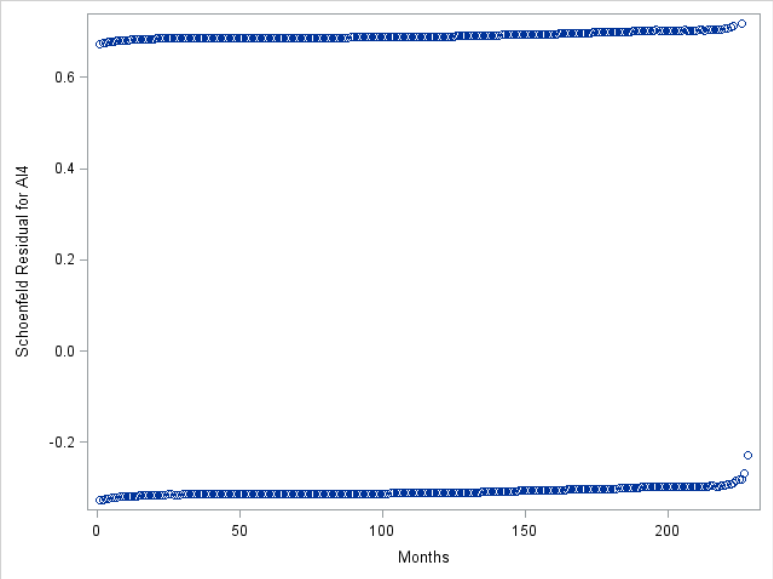

Figure 22. Graphic analysis of proportional hazards assumption using Schoenfeld residuals (Age group 4)

8. APPENDIX F – SAS CODE: CREATING AND ORGANIZING DATASET

```
*************************************************************************/
INPUT: ocr_nodup.sas7bdat
         pccf99.sas7bdat
         pccf01.sas7bdat
         pccf08.sas7bdat
OUTPUT: CENSUS.sas7bdat
*************************************************************************/
options linesize=75 pagesize=55;
footnote1 "Directory: C:\databases\DatabaseUsers\Andrew\sascode\Final Codes";
footnote2 "File: Creating Census.sas7bdat";

libname ocr "C:\databases\central.roru\DSA.data\linkage15";
libname pc "C:\databases\geo\geo_code\pccf.sasdata";
libname SAS "C:\databases\DatabaseUsers\Andrew\sasdata";
************************************************************************
I. Create 2006 Census Data (2003-2009)
************************************************************************;
/* Set PCCF and OCR data to remove duplicates and any cancer cases outside
the 7-year range */

/* Use single link indicator (SLI) to remove duplicate postal codes (PCODE)
in PCCF */

data PC2006;
set pc.pccf08;
if PR = 35;
if SLI = 1;
CODE = 1*DAuid;
run;

/* Select cases diagnosed in 2003-09 from OCR */

data Can2006;
set ocr.Ocr_nodup;
if diag_yr < 2003 then delete;
if diag_yr > 2010 then delete;
PCODE = ORIG_POSTAL_CD;
run;

/* Merge by postal code */

proc sort data=Can2006 out=Cane3;
by PCODE;
run;

proc sort data=PC2006 out=PC2;
by PCODE;
run;

data Cane4;
merge Cane3(in=z) PC2;
by PCODE;
if z=1;
```

```
run;

/* Determine income quintiles in census data */

data DA2;                        /* Import 2006 census data */
set sas.DA2006;
if MHI06 = 0 then MHI06 = .;
run;

proc sort data=DA2 out=DA3;
by Population;
run;

proc univariate data=DA3;
var MHI06;
output out=DA4 pctlpre=P_ pctlpts=0 to 100 by 20;
weight Population;
run;

/* Create median income quintiles */

data sas.MHI2006 (keep=CODE MHI06 q Population);
set DA2;
if MHI06 = 0 or MHI06 = . then q = .;
if MHI06 > 0 or MHI06 < 45000 then q = 1;
if MHI06 >= 45000 and MHI06 < 58157.89474 then q = 2;
if MHI06 >= 58157.89474 and MHI06 < 72500 then q = 3;
if MHI06 >= 72500 and MHI06 < 89062.5 then q = 4;
if MHI06 >= 89062.5 then q = 5;
run;

data DA5;
set sas.MHI2006;
CD = MHI06*116.5/107;
run;

/* Merge census + population with cancer data */

proc sort data=DA5 out=DA6;
by CODE;
run;

proc sort data=Cane4 out=Cane11;
by CODE;
run;

data Cane5;
merge Cane11 (in=e) DA6;
if e=1;
by CODE;
run;

/* Check frequency of SES groups in OCR */

proc freq data=Cane5;
tables q;
run;
```

```
*************************************************************************
II. Create 2001 Census Data (1998-2002)
*************************************************************************;

/* Set PCCF and OCR data to remove duplicates and any cancer cases outside
the 5-year range */

/* Use single link indicator (SLI) to remove duplicate postal codes (PCODE)
in PCCF */

data PC3;
set pc.pccf01;
if PR =35;
if SLI = 1;
CODE = 1*DA_CODE;
run;

/* Select cases diagnosed in 1998-2002 from OCR */

data Can2001;
set Ocr.Ocr_nodup;
if diag_yr < 1998 then delete;
if diag_yr > 2002 then delete;
PCODE = ORIG_POSTAL_CD;
run;

/* Merge by postal code */

proc sort data=Can2001 out=Cand3;
by PCODE;
run;

proc sort data=PC3 out=PC4;
by PCODE;
run;

data Cand4;
merge Cand3 (in=y) PC4;
by PCODE;
if y=1;
run;

/* Determine income quintiles */

proc sort data=sas.EA01 out=EA2; /* Import 2001 Census data */
by CODE;
run;

data EA3;
set EA2;
if _2001_MHI = 0 then _2001_MHI = .;
run;

proc sort data= EA3 out= EA4;
by POP01;
run;
```

```
proc univariate data=EA4;
var _2001_MHI;
output out=EA6 pctlpre=P_ pctlpts=0 to 100 by 20;
weight POP01;
run;

/* Create median income quintiles */

data sas.MHI2001 (keep=CODE _2001_MHI POP01 q);
set EA3;
if _2001_MHI = . then q = .;
if _2001_MHI > 0 and _2001_MHI < 42586 then q = 1;
if _2001_MHI >= 42586 and _2001_MHI < 50000 then q = 2;
if _2001_MHI >= 50000 and _2001_MHI < 60313 then q = 3;
if _2001_MHI >= 60313 and _2001_MHI < 75300 then q = 4;
if _2001_MHI >= 75300 then q = 5;
run;

data EA5;
set sas.MHI2001;
CD = _2001_MHI*116.5/95.4;
run;

/* Merge census + population with cancer data */

proc sort data=EA5 out=EA6;
by CODE;
run;

proc sort data=Cand4 out=Cand5;
by CODE;
run;

data Cand6;
merge Cand5 (in=m) EA6;
if m = 1;
by CODE;
run;

/* Check frequency of SES groups in cancer data */

proc freq data=Cand6;
tables q;
run;

*************************************************************************
III. Create 1996 Census Data (1993-1997)
*************************************************************************;

/* Set PCCF and OCR data to remove duplicates and any cancer cases outside
the 5-year range */

/* Use single link indicator (SLI) to remove duplicate postal codes (PCODE)
in PCCF */

data PC1996;
```

```
set pc.pccf99;
if PR =35;
if SLI = 1;
CODE = 1*EAuid;
run;

/* Select cases diagnosed in 1993-97 from OCR */

data Can1996;
set ocr.Ocr_nodup;
if diag_yr < 1993 then delete;
if diag_yr > 1997 then delete;
PCODE = ORIG_POSTAL_CD;
run;

/* Merge by postal code */

proc sort data=Can1996 out=Cant3;
by PCODE;
run;

proc sort data=PC1996 out=PC2;
by PCODE;
run;

data Cant4;
merge Cant3 (in=x) PC2;
by PCODE;
if x=1;
run;

/* Merge population numbers with median income */

proc sort data=sas.POPS96 out= POP1; /*Import 1996 Ontario Population Data */
by CODE;
run;

proc sort data=sas.EA96 out=EA97; /* Import 1996 Census Data */
by CODE;
run;

data EA98;
merge POP1 EA97;
by CODE;
run;

/* Determine income quintiles */

data EA90;
set EA98;
if IHTMED = 0 then IHTMED = .;

proc sort data=EA98 out=EA99;
by ASPOP;
run;

proc univariate data=EA99;
```

```
var IHTMED;
output out=EA00 pctlpre=P_ pctlpts=0 to 100 by 20;
weight ASPOP;
run;

/* Create median income quintiles */

data sas.MHI1996 (keep=CODE IHTMED ASPOP q);
set EA90;
if IHTMED >= 0 and IHTMED < 34279 then q = 1;
if IHTMED >= 34279 and IHTMED < 42455 then q = 2;
if IHTMED >= 42455 and IHTMED < 51355 then q = 3;
if IHTMED >= 51355 and IHTMED < 62836 then q = 4;
if IHTMED >= 62836 then q = 5;
run;

/* Convert MHI to constant dollars */

data EA02;
set sas.MHI1996;
CD = IHTMED*116.5/87.6;
run;

/* Merge census + population with cancer data */

proc sort data=EA02 out=EA03;
by CODE;
run;

proc sort data=Cant4 out=Cant11;
by CODE;
run;

data Cant5;
merge Cant11 (in=f) EA03;
if f=1;
by CODE;
run;

/* Check frequency of SES groups in OCR */

proc freq data=Cant5;
tables q;
run;

*************************************************************************
IV. Merge all 3 datasets into one census dataset (1993-2009)
*************************************************************************;

/* 2006 Census */

data Cen2006(keep=PERSON_KEY DEATH_DATE_YEAR CAUSE_ORIG_ICD_CD_VERSION
CAUSE_ORIG_ICD_CD CD death_date_key dodx birth_date_key AGE site3 diag_yr
PCODE CODE Population MHI06 q SEX_CD);
set Cane5;
run;
```

```
/* 2001 Census */

data Cen2001 (keep=PERSON_KEY DEATH_DATE_YEAR CAUSE_ORIG_ICD_CD_VERSION
CAUSE_ORIG_ICD_CD CD death_date_key dodx birth_date_key AGE site3 diag_yr
PCODE CODE POP01 _2001_MHI q SEX_CD);
set Cand6;
run;

/* 1996 Census */

data Cen1996 (keep=PERSON_KEY DEATH_DATE_YEAR CAUSE_ORIG_ICD_CD_VERSION
CAUSE_ORIG_ICD_CD CD death_date_key dodx birth_date_key AGE site3 diag_yr
PCODE CODE ASPOP IHTMED q SEX_CD);
set Cant5;
run;

/* Merge Census */

data sas.OCR_SES;
set Cen2006(in=c3) Cen2001(in=c2) Cen1996(in=c1);
if c1=1 then YEAR=1996;
if c2=1 then YEAR=2001;
if c3=1 then YEAR=2006;
run;

/* Clean merged census data */

data Combined3;
set sas.OCR_SES;
if CD = 0 then CD = '.';
if AGE > 110 then AGE = '.';
run;

/* Convert age to age group */

proc univariate data=Combined3;
var AGE;
output out=Combined4 pctlpre=P_ pctlpts=0 to 100 by 20;
run;

data Combined5;
set Combined3;
if AGE < 50 then AI = 1;
if AGE >= 50 and AGE < 60 then AI=2;
if AGE >=60 and AGE < 70 then AI=3;
if AGE >= 70 and AGE < 80 then AI=4;
if AGE >= 80 then AI=5;
run;

/* Convert "Year of Diagnosis" into a numeric variable */

data Combined6;
set Combined5;
DYEAR = 1*diag_yr;
run;

/* Drop unnecessary variables */
```

```
data Combined7 (keep = DEATH_DATE_YEAR death_date_key dodx site3 q DYEAR AI
AGE SEX_CD CAUSE_ORIG_ICD_CD_VERSION CAUSE_ORIG_ICD_CD CD);
set Combined6;
run;

*************************************************************************
V. Include survival variables
*************************************************************************;

/*Convert Death Year to a numeric variable */

data Surv;
set Combined7;
if DEATH_DATE_YEAR = . then fstat=0;
else fstat=1;
run;

data Surv2;
set Surv;
if CAUSE_ORIG_ICD_CD_VERSION = 10 then ICD10 = CAUSE_ORIG_ICD_CD;
if CAUSE_ORIG_ICD_CD_VERSION = 9 then ICD9 = CAUSE_ORIG_ICD_CD;
if "C00" > ICD10 or ICD10 > "C999" then ICD1=0;
else ICD1=1;
if "1400" > ICD9 or "2089" < ICD9 then CAN=0;
else CAN = 1;
if ICD1 = 1 or CAN=1 then dstat=1;
if fstat=0 then dstat=0;
if fstat=1 and ICD1=0 and CAN=0 then dstat=2;
run;

/*Calculate the difference between the date of diagnosis and date of death */

data Surv3;
set Surv2;
if DEATH_DATE_YEAR = . or DEATH_DATE_YEAR > 2013 then do;
Days = "31DEC2013"D - dodx;
fstat=0;
end;
else Days = death_date_key - dodx;
run;

/*Make sex a numeric variable and clean data set further */

data Surv4;
set Surv3;
if SEX_CD = 'M' then SEX = 1;
else SEX = 0;
run;

data Surv5;
set Surv4;
if Days = '.' then delete;
if Days < 0 then delete;
run;

/* Replace event status with competing event status, when appropriate */
```

```
data sas.CENSUS;
set Surv5;
if DEATH_DATE_YEAR > 2011 or DEATH_DATE_YEAR = . then do;
DAYS2 = "31DEC2011"D - dodx;
dstat=0;
end;
else DAYS2=DAYS;
run;

************************************END************************************;
```

9. APPENDIX G – SAS CODE: ANALYZING DATASET

```
*********************************************************************/
INPUT: CENSUS.sas7bdat
OUTPUT:
*********************************************************************/
options linesize=75 pagesize=55;
footnote1 "Directory: C:\databases\geo\geo_code\pccf.sasdata";

libname SAS "C:\databases\DatabaseUsers\Andrew\sasdata";
*********************************************************************
I. Set-up covariates and follow-up months
*********************************************************************;

/* Check Study Population */

proc univariate data=sas.CENSUS;
var AGE;
run;

proc sort data=sas.CENSUS out=GEN;
by SEX;
run;

proc freq data=GEN;
tables SEX;
run;

proc sort data=sas.CENSUS out=MHI;
by q;
run;

proc univariate data=MHI;
var CD;
by q;
run;

proc sort data=sas.CENSUS out=DIAGN;
by DYEAR;
run;
```

```
proc freq data=DIAGN;
tables DYEAR;
run;

proc sort data=sas.CENSUS out=SES;
by q;
run;

proc freq data=SES;
tables q / Missing;
run;

proc freq data=sas.CENSUS;
tables AI;
run;


/* Convert days of follow-up to months of follow-up */

data Mon;
set sas.CENSUS;
Months=ceil(Days/30.4);
Months2=ceil(Days2/30.4);
run;

*************************************************************************
II. Run Fine-Gray Subdistribution Hazards model and check PH assumption (All
sites)
*************************************************************************;

/* Check if PH assumption holds */

proc phreg data=Mon alpha=0.05;
class q SEX AI;
model Months*dstat(0) = q|DYEAR SEX AI / eventcode=1;
output out=SCH ressch=q1 q2 q3 q4 DIAG a b c d GEND AG1 AG2 AG3 AG4;
run;

proc sgplot data=SCH;
loess x=Days y=q1;
run;

proc sgplot data=SCH;
loess x=Days y=q2;
run;

proc sgplot data=SCH;
loess x=Days y=q3;
run;

proc sgplot data=SCH;
loess x=Days y=q4;
run;

proc sgplot data=SCH;
loess x=Days y=DIAG;
run;
```

```
proc sgplot data=SCH;
loess x=Days y=a;
run;

proc sgplot data=SCH;
loess x=Days y=b;
run;

proc sgplot data=SCH;
loess x=Days y=c;
run;

proc sgplot data=SCH;
loess x=Days y=d;
run;

proc sgplot data=SCH;
loess x=Days y=GEND;
run;

proc sgplot data=SCH;
loess x=Days y=AG1;
run;

proc sgplot data=SCH;
loess x=Days y=AG2;
run;

proc sgplot data=SCH;
loess x=Days y=AG3;
run;

proc sgplot data=SCH;
loess x=Days y=AG4;
run;


/* Run Kaplan-Meier estimate to calculate 5-year overall survival, stratified
by SES */

proc lifetest data=sas.CENSUS method=km intervals=(0 to 1825 by 1)
plots=survival(cb) conftype=loglog
alpha=0.05;
strata q;
time Months*fstat(0);
run;

/* Run Kaplan-Meier estimate to calculate 5-year overall survival, stratified
by year of diagnosis */

proc lifetest data=sas.CENSUS method=km intervals=(0 to 1825 by 1)
plots=survival(cb) conftype=loglog
alpha=0.05;
strata DYEAR;
time Months*fstat(0);
run;
```

```
/* Calculate cumulative incidence function, grouped by SES */

%CIF(data=Mon, TIME=Months2, STATUS=dstat,EVENT=1, CENSORED=0, GROUP=q,
ALPHA=0.05, OPTIONS=plotcl, out=SES);

/* Calculate cancer-specific survival */

data CIF2;
set SESplot;
SURV = 1 - CIF;
run;

/* Calculate cumulative incidence function, grouped by year of diagnosis */

%CIF(data=Mon, TIME=Months2, STATUS=dstat,EVENT=1, CENSORED=0, GROUP=DYEAR,
ALPHA=0.05, OPTIONS=plotcl, out=DIAG);

/* Calculate cancer-specific survival */

data CIF2;
set DIAGplot;
SURV = 1 - CIF;
run;

/* Calculate cumulative incidence function, grouped by SES and stratified by
year of diagnosis */

%CIF(data=Mon, TIME=Months2, STATUS=dstat,EVENT=1, CENSORED=0, GROUP=q,
STRATA=DYEAR, ALPHA=0.05, OPTIONS=plotcl, out=CIF);

/* Calculate cancer-specific survival */

data CIF2;
set CIFplot;
SURV = 1 - CIF;
run;

/* Run Fine-Gray Model for overall model */

proc phreg data=Mon alpha=0.05;
class q SEX AI;
model Months2*dstat(0) = q|DYEAR SEX AI / eventcode=1;
hazardratio DYEAR;
run;

*************************************************************************
II. Run Fine-Gray Subdistribution Hazards Regression (Site-Specific)
*************************************************************************;

/* Includes Breast (1), Lung (2), Head & Neck (3), & Colorectal (4) */

data Sites;
set Mon;
if site3 = 'C50' then Site = 1;
if site3 = 'C34' then Site = 2;
```

```
if site3 = 'C00' or site3 = 'C01' or site3 = 'C02' or site3 = 'C03' or site3
= 'C04'
or site3 = 'C05' or site3 = 'C06' or site3 = 'C07' or site3 = 'C08' or site3
= 'C09'
or site3 = 'C10' or site3 = 'C11' or site3 = 'C12' or site3 = 'C13' or site3
= 'C14'
or site3 = 'C30' or site3 = 'C31' or site3 = 'C32' or site3 = 'C33'
then Site=3;
if site3 = 'C18' or site3 = 'C19' or site3 = 'C20' or site3 = 'C21' or site3
= 'C26'
then Site = 4;
run;

data Breast;
set Sites;
if Site=1;
run;

data Lung;
set Sites;
if Site=2;
run;

data HN;
set Sites;
if Site=3;
run;

data Colorectal;
set Sites;
if Site=4;
run;

/* Run Kaplan-Meier estimate to calculate 5-year overall survival, stratified
by SES */

proc lifetest data=Sites method=km intervals=(0 to 1825 by 1)
plots=survival(cb) conftype=loglog
alpha=0.05;
strata q;
by Site;
time Months*fstat(0);
run;

/* Run Kaplan-Meier estimate to calculate 5-year overall survival, stratified
by year of diagnosis */

proc lifetest data=Sites method=km intervals=(0 to 1825 by 1)
plots=survival(cb) conftype=loglog
alpha=0.05;
strata DYEAR;
by Site;
time Months*fstat(0);
run;

/* Calculate cumulative incidence functions for the specified cancer sites,
grouped by SES and stratified by year of diagnosis */
```

```
%CIF(data=Breast, TIME=Months2, STATUS=dstat,EVENT=1, CENSORED=0, GROUP=q,
STRATA=DYEAR, ALPHA=0.05, OPTIONS=plotcl, out=Breast1);
%CIF(data=Lung, TIME=Months2, STATUS=dstat,EVENT=1, CENSORED=0, GROUP=q,
STRATA=DYEAR, ALPHA=0.05, OPTIONS=plotcl, out=Lung1);
%CIF(data=HN, TIME=Months2, STATUS=dstat,EVENT=1, CENSORED=0, GROUP=q,
STRATA=DYEAR, ALPHA=0.05, OPTIONS=plotcl, out=HN1);
%CIF(data=Colorectal, TIME=Months2, STATUS=dstat,EVENT=1, CENSORED=0,
GROUP=q, STRATA=DYEAR, ALPHA=0.05, OPTIONS=plotcl, out=Colorectal1);

/* Calculate cancer-specific survival */

data Breast2;
set Breast1plot;
SURV = 1 - CIF;
run;


data Lung2;
set Lung1plot;
SURV = 1 - CIF;
run;

data HN2;
set HN1plot;
SURV = 1 - CIF;
run;

data Colorectal2;
set Colorectal1plot;
SURV = 1 - CIF;
run;

/* Run Fine-Gray Model for all cancers sites */

proc sort data=Sites;
by Site;
run;

proc phreg data=Sites alpha=0.05;
class q SEX AI;
model Months2*dstat(0) = q|DYEAR SEX AI / eventcode=1;
hazardratio DYEAR;
by Site;
run;

/* Assess missing data */

data MISSING;
set Mon;
if q = . then P = 1;
else P = 0;
run;

proc lifetest data=MISSING method=km intervals=(0 to 1825 by 1)
plots=survival(cb) conftype=loglog alpha=0.05;
strata P;
```

```
time Months*fstat(0);
run;

**********************************END**********************************;
```