Effects of the Length of Central Cancer Registry Operations on Identification of Subsequent Cancers and on Survival Estimates

Baozhen Qiao, Maria J. Schymura, Amy R. Kahn, New York State Cancer Registry, Albany, NY

INTRODUCTION

Background: The common practice in population-based cancer survival analyses is to include only the first primary cancer diagnosis and exclude any subsequent primaries for an individual. However, several recent studies have brought this practice into question and have recommended including all primary cancers. One of the arguments against using only the first primary cancer is that the reference date of a cancer registry's operations affects the registry's ability to identify earlier cancers. A registry with an earlier reference date is more likely to identify previous cancer diagnoses compared to a new registry. Therefore, a tumor that is in fact a subsequent primary might be erroneously treated as a first primary if the registry is not in operation long enough. Based on the first primary rule, such a tumor would be included in the survival calculation in a registry with a later reference date, but would be excluded in an older registry. This selection bias would cause survival comparisons across registries with different reference dates to be less reliable.

Objectives: In this study, we used a theoretical approach to evaluate how and to what extent the length of central cancer registry operations affects the ascertainment of subsequent cancers and consequently on survival estimates, with or without including subsequent cancers.

METHODS

Source of data:

- All invasive cancer cases (including *in situ* bladder cancers) diagnosed from 2001 to 2010 in New York State were used for the study.

- Sequence number central (NAACCR data item #380) was used to define first vs. subsequent primary cancers.

- A set of three sequence numbers was created for each tumor. The first sequence number was the originally coded and stored value in the database, derived according to the true reference year, 1976, for the New York State Cancer Registry (NYSCR). The second and third sequence numbers were derived based on two hypothetical reference dates, the year of 1986 or 1996, respectively.

- Cancer cases with a sequence number value of 00 or 01 were classified as the first primary cancer; while cancer cases with a sequence number value of 02 or higher were classified as subsequent primary cancers. The status of first vs. subsequent primary for a particular tumor might change depending on the assumed reference year. Table 1 shows how the sequence number central and first/subsequent primary status were assigned based on different reference years.

Table 1. Example showing how the sequence number central and first/subsequent primary cancer status were assigned for a patient with three primary cancer diagnoses. Only Tumor 3 was the interest of the study.

Primary Cancer	Year of Diagnosis	Sequence Number Central			First Primary (F) or Subsequent Primary (S)		
		Reference Year: 1976	Reference Year: 1986	Reference Year: 1996	Reference Year: 1976	Reference Year: 1986	Reference Year: 1996
Tumor 1	1984	01	-	-	F	-	-
Tumor 2	1994	02	01	-	S	F	-
Tumor 3	**2004**	**03**	**02**	**00**	**S**	**S**	**F**

Subsequent primary cancers:

- Percents of subsequent primaries classified based on the different reference years were calculated for all sites combined and 23 major cancer groups.

- Median age at diagnosis was also calculated separately for cancers classified as first primaries and cancers classified as subsequent primaries.

- Percent of subsequent primary cancers was evaluated by year of diagnosis and among different age groups (<44, 45-54, 55-64, 65-74, 75+).

Survival analysis:

- Five-year relative survival rates were calculated for all sites combined and for 23 major cancer groups using four different selection criteria: 1) only first primary cancer, determined based on the reference year of 1976; 2) only first primary cancer, determined based on reference year of 1986; 3) only first primary cancer, determined based on the reference year of 1996; and 4) all primary cancers.

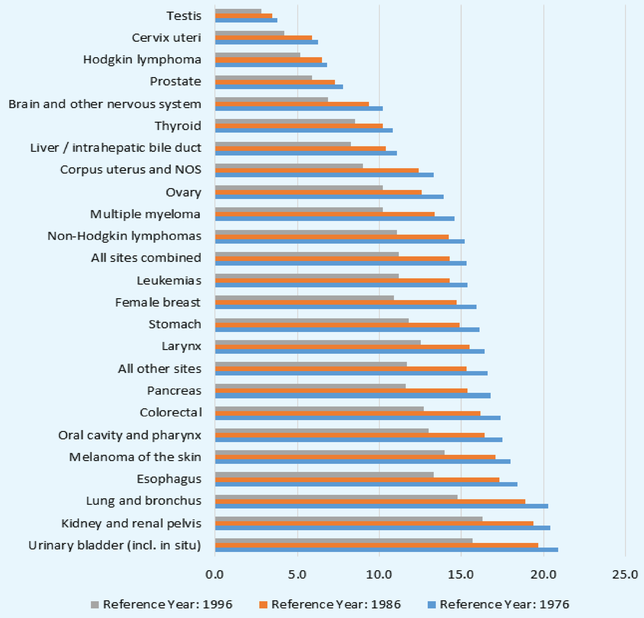
- The absolute survival difference was used to measure the magnitude of the effect on survival rates under different selection scenarios.

RESULTS

- The percent of tumors classified as subsequent cancers decreased when the registry operation time was shortened. Using the reference year of 1976, 15.3% of cases were classified as subsequent cancer diagnoses, while the percent of subsequent cancers decreased to 14.3% and 11.2%, respectively, for the reference years 1986 and 1996.

- The percent of tumors classified as subsequent cancers varied considerably among cancer sites, ranging from 3.8% for testis cancer to 20.9% for urinary bladder cancer for the reference year of 1976, and from 2.8% for testis cancer to 16.3% for kidney and renal pelvis cancer for the reference year of 1996 (Figure 1).

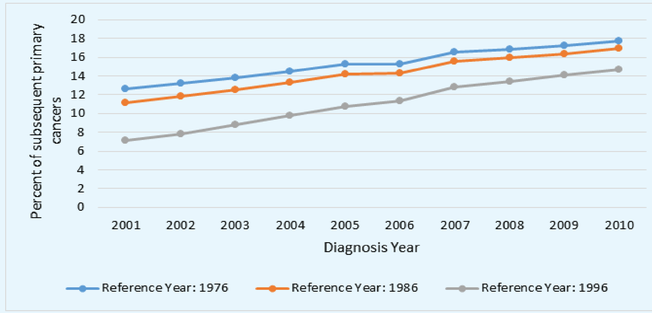
Figure 1. Percent of subsequent primary cancers by cancer site under three different registry reference year scenarios



- As expected, median age at diagnosis was always older for the subsequent cancers compared to the first primary cancers.

- The percent of tumors classified as subsequent cancers increased over time regardless of reference year. However, this increase was more rapid for a later reference date (Figure 2).

Figure 2. Trends in the percent of subsequent primary cancers under three different registry reference year scenarios



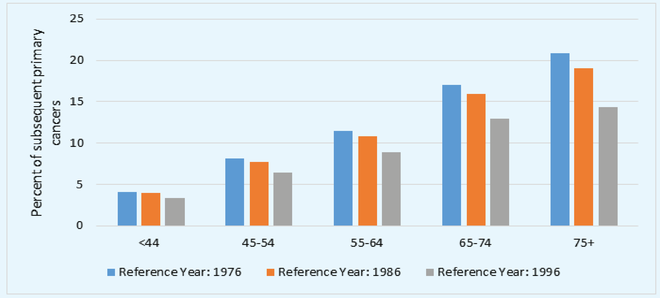
Acknowledgement

This work was funded in part by CDC Cooperative Agreement 5U58DP000783 awarded to the New York State Department of Health.

RESULTS - continued

- Subsequent cancers were more likely to be diagnosed among older patients. The effect of registry reference date on percent of subsequent cancers was also more profound for the older age group (Figure 3).

Figure 3. Percent of subsequent priamry cancers by age group under three different registry reference year scenarios



- When only the first primary cancer was included in the survival calculations, shorter registry operation time was associated with slightly lower 5-year relative survival estimates; however, the maximum survival difference did not exceed one percent for the scenario with a 20-year difference in registry operation length (Table 2).

- When all primary cancers were included, survival estimates decreased for almost all types of cancers. These decreases were the greatest for the earliest reference year. The magnitude of the survival decrease also varied by cancer site (Table 2).

Table 2. Five-year relative survival rates by cancer site, with inclusion of only first primary cancer vs. all primary cancers under three different registry reference year scenarios

Cancer Sites	Only First Primary (sequence number 0 or 1)			All Primaries (B)	Survival Difference			
	Reference Year : 1976 (A1)	Reference Year: 1986 (A2)	Reference Year: 1996 (A3)		A3-A1	B-A1	B-A2	B-A3
All sites combined	65.9	65.7	65.3	64.9	-0.6	-1.0	-0.8	-0.4
Oral cavity and pharynx	60.7	60.6	60.2	58.7	-0.5	-2.0	-1.9	-1.5
Esophagus	18.0	18.0	18.0	17.6	0.0	-0.4	-0.4	-0.4
Stomach	33.5	33.3	32.9	32.2	-0.6	-1.3	-1.1	-0.7
Colorectal	64.1	63.9	63.7	62.8	-0.4	-1.3	-1.1	-0.9
Liver / intrahepatic bile duct	20.5	20.5	20.3	20.0	-0.2	-0.5	-0.5	-0.3
Pancreas	9.0	9.0	8.8	8.7	-0.2	-0.3	-0.3	-0.1
Larynx	62.4	62.4	62.0	60.3	-0.4	-2.1	-2.1	-1.7
Lung and bronchus	20.1	20.1	20.2	20.6	0.1	0.5	0.5	0.4
Melanoma of the skin	90.1	90.0	89.6	88.5	-0.5	-1.6	-1.5	-1.1
Female breast	88.7	88.6	88.5	88.0	-0.2	-0.7	-0.6	-0.5
Cervix uteri	69.1	69.0	68.7	67.8	-0.4	-1.3	-1.2	-0.9
Corpus uterus and NOS	79.9	79.8	79.3	78.2	-0.6	-1.7	-1.6	-1.1
Ovary	46.9	46.7	46.4	46.6	-0.5	-0.3	-0.1	0.2
Prostate	97.6	97.6	97.4	96.4	-0.2	-1.2	-1.2	-1.0
Testis	96.1	96.1	96.2	96.0	0.1	-0.1	-0.1	-0.2
Urinary bladder (incl. in situ)	78.5	78.3	77.7	75.8	-0.8	-2.7	-2.5	-1.9
Kidney and renal pelvis	74.0	73.9	73.5	72.2	-0.5	-1.8	-1.7	-1.3
Brain and other nervous system	37.4	37.2	36.5	34.9	-0.9	-2.5	-2.3	-1.6
Thyroid	98.1	98.0	97.8	96.8	-0.3	-1.3	-1.2	-1.0
Hodgkin lymphoma	84.3	84.3	83.9	82.5	-0.4	-1.8	-1.8	-1.4
Non-Hodgkin lymphomas	68.5	68.3	67.8	66.7	-0.7	-1.8	-1.6	-1.1
Multiple myeloma	46.9	46.8	46.4	45.4	-0.5	-1.5	-1.4	-1.0
Leukemia	56.3	56.1	55.5	53.3	-0.8	-3.0	-2.8	-2.2

CONCLUSIONS

The current study showed that the length of registry operations was positively associated with the percentage of tumors classified as subsequent cancer diagnoses.

The overall effect of sequence number misclassification and selection bias on survival estimates was small if the length of registry operations differed by less than 20 years. However, this survival bias could increase when the difference of operation length between registries becomes larger.

Although the survival estimates did not vary greatly between our different reference year assumptions, estimates based only on presumed first cancers does introduce misclassification bias that would be eliminated if the estimates included all cancers. Therefore, inclusion of all primary cancers is recommended.