Impact of Cancer Incidence Reporting Delay on Population-Based Survival Analysis

[2]X Dong, [2]Y Ren, [2]K Zhang, [1]R Wilson

[1]Division of Cancer Prevention and Control, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, Atlanta, GA

[2]ICF International, Fairfax, VA

Introduction:

1. Relative survival methodology with presumed-alive assumption has been adopted by NPCR and NAACCR for population-based cancer survival analysis. However, Pinhero *et. al.* (2014) and Johnson *et. al.* (2010) have observed overestimation biases with use of these methods.
2. The survival time of a cancer patient is directly determined by date of diagnosis and date of last contact. Any reporting issues in these variables, such as incidence reporting delay, may cause biases in survival outcomes.
3. The purpose of this study is to assess the impact of incidence reporting delay to identify and quantify potential biases in population-based survival analysis with presumed-alive assumption.

Study Data:

1. Incidence reporting delay was calculated from NPCR data submissions 2001-2015 from 16 registries who met NPCR data quality standards, demonstrated consistent patient IDs between submissions, and performed NDI linkages.
2. The foundation population included the malignant cases from the 16 registries in the NPCR 2015 data submission with age between 0-99.
3. Two study populations were formed from the foundation population. a) Population 1: Cases diagnosed between 2000-2006 with follow-up to 2006; b) Population 2: Cases diagnosed between 2001-2007 with follow-up to 2007.

Method:

1. To isolate the effect of incidence reporting delay, the study populations used values of sex, race, date of diagnosis, date of last contact, vital status, age of diagnosis from the foundation population throughout the study.
2. Population 1 and Population 2 were analyzed by traversing the initial cohort through individual submission from period 2009-2015 and 2010-2015, respectively. Cases were selected into each submission year by incidence reporting delays. The vital status and date of last contact were recalculated and reassigned for each submission year. The only varying factor each year is the addition of newly found delayed cases.
3. Used relative survival methodology: Ederer II method for the expected survival with NCHS 2000-2010 national level life tables; actuarial method for the observed survival; used customized NPCR relative survival SAS tool for estimation. The survival outcomes from Submission 2015 are considered stable. Analyses were done by site, race, sex, and age. Only site level is reported in this presentation.

Results:

Submission Year	Year of Diagnosis 1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
2001	335788	0	0	0	0	0	0	0	0	0
2002	36148	346478	0	0	0	0	0	0	0	0
2003	6091	18036	360319	0	0	0	0	0	0	0
2004	9946	8503	15342	376830	0	0	0	0	0	0
2005	2117	3153	4913	10428	380203	0	0	0	0	0
2006	2582	2702	3642	5057	10413	378418	0	0	0	0
2007	1304	1564	2574	3072	4952	10416	386365	0	0	0
2008	687	892	1349	1724	2390	4123	8885	392593	0	0
2009	543	640	1007	1168	1376	1964	3178	6568	402634	0
2010	513	567	1009	928	1097	1424	2145	3550	8859	415135
2011	434	571	580	586	731	983	1484	1927	3256	8791
2012	1363	965	949	964	968	1074	1381	1886	2525	3740
2013	359	464	544	621	650	708	1008	1339	1518	2156
2014	368	364	420	821	988	751	791	1137	1793	2456
2015	327	307	414	457	479	552	652	902	874	1184
Total	**398570**	**385206**	**393062**	**402656**	**404247**	**400413**	**405889**	**409902**	**421459**	**433462**
				Initial population1				**Initial population2**		

Table 1: Incidence reporting delays between 1998 and 2007 in NPCR data submission 2001-2015 from 16 registries. The number at the top of each column represents the cases reported without delay in a 24-month reporting window of a diagnosis year. The initial cohorts of Population 1 and 2 are as color coded. The cases from submission 2010-2015 are the reporting delayed cases sequentially added to the starting cohorts.

Cancer site	Submission year	Cases	5-year relative survival	Standard error	Incidence reporting delay induced underestimation of survival
Colon & rectum	2009	288846	62.40%	0.15%	*-0.29%
Colon & rectum	2010	290307	62.50%	0.15%	-0.19%
Colon & rectum	2011	291042	62.56%	0.15%	-0.13%
Colon & rectum	2012	291760	62.62%	0.15%	-0.08%
Colon & rectum	2013	292224	62.66%	0.15%	-0.03%
Colon & rectum	2014	292753	62.67%	0.15%	-0.02%
Colon & rectum	2015	292999	62.69%	0.15%	0.00%
Leukemia	2009	67466	**47.78%	0.28%	-2.93%
Leukemia	2010	68812	48.59%	0.27%	-2.11%
Leukemia	2011	69758	49.20%	0.27%	-1.51%
Leukemia	2012	70501	49.70%	0.27%	-1.01%
Leukemia	2013	71092	50.08%	0.27%	*-0.63%
Leukemia	2014	71602	50.39%	0.27%	-0.32%
Leukemia	2015	72093	50.71%	0.27%	0.00%

* The submission year that the survival was no longer significantly different from those of Submission 2015. From this submission year and beyond, the survival outcomes are considered stablized in each site.

** The cumulative relative survival of Leukemia in Table 2 are statistically significant different from their correspondents in Table 3.

Table 2: The effects of incidence reporting delay on 5-year relative survival outcomes of Population 1 are demonstrated with colon & rectum and leukemia. The underestimation of survival is the net difference between survival estimates of earlier submission years and Submission 2015. The reporting delay has no effect on colon & rectum, but cause significant underestimation of survivals in leukemia. Leukemia survival estimates stabilized after 4 years of subsequent submissions.

Cancer site	Submission year	Cases	5-year relative survival	Standard error	Incidence reporting delay induced underestimation of survival
Colon & rectum	2010	285928	62.69%	0.15%	*-0.23%
Colon & rectum	2011	287268	62.77%	0.15%	-0.16%
Colon & rectum	2012	288148	62.84%	0.15%	-0.09%
Colon & rectum	2013	288707	62.88%	0.15%	-0.04%
Colon & rectum	2014	289352	62.90%	0.15%	-0.02%
Colon & rectum	2015	289647	62.92%	0.15%	0.00%
Leukemia	2010	69120	**49.50%	0.27%	-2.48%
Leukemia	2011	70451	50.22%	0.27%	-1.76%
Leukemia	2012	71391	50.80%	0.27%	-1.18%
Leukemia	2013	72117	51.24%	0.27%	-0.74%
Leukemia	2014	72721	51.60%	0.27%	*-0.38%
Leukemia	2015	73310	51.98%	0.27%	0.00%

Table 3: The effects of incidence reporting delay on 5-year relative survival outcomes of Population 2 are demonstrated with colon & rectum and leukemia. The underestimation of survival is the net difference between survival estimates of earlier submission years and Submission 2015. The reporting delay has no effect on colon & rectum, but cause significant underestimation of survivals in leukemia. Leukemia survival estimates stabilized after 4 years of subsequent submissions.

Study Population	Cancer site	Max observed underestimation	Extra years of submission to reach stablization
2000-2006	All sites Combined	-0.47%	4
	Lung & bronchus	-0.30%	1
	Brain & other nervous system	-0.97%	1
	Non-Hodgkin lymphoma	-0.81%	2
	Myeloma	-1.88%	2
	Leukemia	-2.93%	4
2001-2007	All sites Combined	-0.40%	3
	Lung & bronchus	-0.27%	1
	Brain & other nervous system	-0.74%	1
	Non-Hodgkin lymphoma	-0.67%	1
	Myeloma	-1.75%	2
	Leukemia	-2.48%	4

Table 4: Six cancer sites whose survival estimates are significantly affected by the incidence reporting delay in Population 1 and 2. Population1, included diagnosis year 2000 which has more reporting delay than later diagnosis years, tends to have slightly greater biases than population2. These sites need additional submissions to achieve stabilized survival estimates in both study populations, especially all-sites-combined, myeloma, and leukemia.

Results:

1. In common scenarios where the latest data submission and most recent diagnosis period are used to examine incidence, reporting delays likely contribute some underestimation of cancer survival. We found this underestimation to be statistically significant in all-sites-combined, brain & other CNS, leukemia, lung & bronchus, myeloma, and non-Hodgkin lymphoma. Myeloma and leukemia have the greatest underestimation in survival and need 2-4 years to overcome.
2. The study population selection has a significant impact on survival outcomes. Early diagnosis years in NPCR data, such as 1998-2000, have severe incidence reporting delays. By avoiding 2000, leukemia survival estimates in Population2 are higher than those from Population1, as large as 1.27% in net difference; and these survival estimates are statistically significantly different. Similar results are observed for myeloma and non-Hodgkin lymphoma, where high incidence reporting delays are expected.

Conclusions:

1. For a mature population-based cancer surveillance system, assuming incidence reporting delay is the single source of bias, survival is under-estimated with the presumed-alive assumption.
2. Recent NPCR data have less, yet more stable, incidence reporting delay. Using diagnosis periods that avoid early NPCR data in the study population may reduce the underestimation bias; and may also significantly increase the survival estimates for cancer sites experiencing severe incidence reporting delay.
3. The current NPCR survival database included diagnosis year 2001 to (the most recent diagnosis year - 1) and corrects 3 less impacted cancer sites. However, for all-sites-combined, myeloma, and leukemia, the period is not enough to mitigate their underestimation.
4. Because of the short history of NPCR data, the study can't evaluate more recent diagnosis years. It would be beneficial to conduct an annual assessment of such biases with each new data submission.
5. It is necessary to continue to identify and quantify other sources of biases that may contribute to the overestimation biases reported in the literature.

Contact:

Xing Dong: xdong@icfi.com

Reda Wilson: dfo8@cdc.gov