Modeling Reporting Delay in the NPCR data

[2]X Dong, [2]K Zhang, [1]T Thompson, [1]Q Lin, [2]Y Ren, [1]R Wilson, [1]C Eheman

[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. National Program of Cancer Registries (NPCR) of CDC funds 48 central cancer registries in the U.S. and its territories.
2. Historically, the cancer incidences reported in the NPCR are not adjusted for reporting delay and reporting errors.
3. Reporting delay happens when a case reports to CDC after its designated 24-month reporting window. Reporting error is removing a case from a subpopulation based on a set of predefined criteria.
4. The NPCR reporting delay model is to explore a better modeling approach to estimate reporting delay distributions that suits the characteristics of the NPCR data.
5. The resulting predicted reporting delay distributions can be used to accurately describe the cancer incidences in the NPCR data.

Methods

❑ Data:

1. Reporting delays and errors defined similarly to what Midthune et. al. described in 2005.
2. The NPCR data used in the model development included 23 central registries and submission 2000 to 2010
3. Variables used to define reporting delays and errors were registry, primary cancer site, sex (male and female), race (white, black and other), age group (18 groups, 0-4, 5-9, … , ≥85), and year of diagnosis (1998 to 2007).
4. The NPCR model was initially developed with melanoma (white only) incidences in the NPCR data. It was later applied to all other cancer sites. Additional sites reported here are prostate (male only), breast (female only), lung and colon & rectum.

Year of Submission	Year of Diagnosis									
	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
2001	16161									
2002	+2587 -1722	17386								
2003	+1376 -871	+2089 -587	19358							
2004	+477 -71	+689 -80	+1280 -101	21707						
2005	+269 -61	+290 -107	+355 -130	+1110 -209	23407					
2006	+142 -66	+130 -96	+209 -140	+406 -215	+968 -228	22690				
2007	+128 -59	+147 -50	+276 -54	+400 -76	+465 -89	+953 -173	24868			
2008	+110 -36	+151 -38	+210 -51	+221 -77	+450 -86	+538 -108	+1502 -168	27532		
2009	+64 -15	+67 -11	+88 -15	+113 -24	+123 -28	+177 -29	+297 -67	+565 -107	28166	
2010	+92 -18	+101 -35	+150 -30	+238 -37	+344 -76	+345 -58	+437 -51	+701 -95	+1251 -135	28166
Initial	16161	17386	19358	21707	23407	22690	24868	27532	28166	28166
Total Delay	5245	3664	2568	2488	2350	2013	2236	1266	1251	0
Total Error	2919	1004	521	638	507	368	286	202	135	0
Net Total	18487	20046	21405	23557	25250	24335	26818	28596	29282	28166

Table 1: The reporting delays and errors patterns of melanoma in 23 central registries in the NPCR data. Note: Table 1 structured similarly to the table published by Midthune et. al. in 2005. The net total is the cumulated melanoma incidences by submission 2010 (Net total = Initial + (Total Delay) – (Total Error)).

❑ Methodology: the NPCR net case truncated delay model (fixed effect model)

1. The NPCR delay model assumed that the net differences between reporting delays and errors are independent Poisson distributions.
2. The net differences can also be considered as binomial distributions conditional on the total reported incidences.
3. The conditional reporting delay probabilities can be estimated with the generalized linear model under the binomial distribution assumption with complementary-log-log link function.
4. Predictor: indicators of delay time period, such as 2, 3, … , or 11 years of delay in reporting
5. Variances of predicted delay distributions can be estimated as Brookmeyer et. al. proposed in 1990 with Greenwood's formula.

Results and Discussions

❑ Comparing the NPCR fixed effect delay model with the NCI fixed effect delay model using melanoma

	NPCR Fixed Effect Model					NCI Fixed Effect Model				
Year of Diagnosis	Melanoma	Breast	Prostate	Lung	Colon & Rectum	Melanoma	Breast	Prostate	Lung	Colon & Rectum
1998	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1999	0.9961	0.9974	0.9962	0.9999	0.9993	0.9972	0.9985	0.9974	0.9999	0.9995
2000	0.9932	0.9955	0.9935	0.9995	0.9985	0.9939	0.9967	0.9942	0.9996	0.9988
2001	0.9891	0.9937	0.9894	0.9990	0.9975	0.9898	0.9946	0.9903	0.9992	0.9979
2002	0.9838	0.9911	0.9847	0.9982	0.9960	0.9848	0.9921	0.9855	0.9985	0.9965
2003	0.9776	0.9882	0.9787	0.9973	0.9944	0.9787	0.9891	0.9797	0.9975	0.9947
2004	0.9703	0.9845	0.9717	0.9954	0.9918	0.9713	0.9856	0.9726	0.9961	0.9923
2005	0.9588	0.9785	0.9610	0.9905	0.9864	0.9597	0.9796	0.9619	0.9911	0.9869
2006	0.9424	0.9710	0.9457	0.9809	0.9781	0.9432	0.9720	0.9466	0.9816	0.9786
2007	0.9025	0.9519	0.9096	0.9506	0.9554	0.9034	0.9529	0.9104	0.9513	0.9559

Table 2: Comparisons of the predicted delay distributions between the NCPR fixed effect delay model and the NCI fixed effect delay model of 5 cancer sites in submission 2010. The predicted reporting delay distributions between the NPCR and NCI models are similar to each other in 5 cancer sites tested. There is resemblance between the predicted delay distributions of melanoma and prostate cancer. It is noticeable that breast cancer has less reporting delays than prostate cancer in the NPCR data. The upward reporting delay adjustment to the incidences diagnosed in 2007 and reported in 2010 ranges from 4.7% to 10.8% in 5 cancer sites tested.

	NPCR Fixed Effect Model			NCI Fixed Effect Model		
Year of Diagnosis	Crude	Delay Adjusted	Delay Adjusted Age Standardized	Crude	Delay Adjusted	Delay Adjusted Age Standardized
1998	21.41	21.41	20.36	21.41	21.41	20.36
1999	23.05	23.14	21.91	23.05	23.12	21.87
2000	24.46	24.63	23.21	24.46	24.61	23.19
2001	26.75	27.04	25.39	26.75	27.02	25.37
2002	28.48	28.95	27.04	28.48	28.92	27.00
2003	27.30	27.93	25.86	27.30	27.88	25.82
2004	29.89	30.81	28.40	29.89	30.76	28.35
2005	31.67	33.03	30.28	31.67	32.97	30.22
2006	32.17	34.14	31.03	32.17	34.07	30.96
2007	30.71	34.03	30.70	30.71	34.00	30.67

Table 3: Comparisons of the effects of delay adjustment on melanoma incidence rates between the NPCR fixed effect delay model and the NCI fixed effect delay model. There is close similarity of the delay adjusted melanoma incidence rates between the two models. By incorporating reporting delay effects, both models level the apparent drop of crude melanoma incidence rates between the diagnosis year 2006 and 2007. The effects of reporting delay adjustment can be seen more obviously in Figure 1.

❑ Results from the NPCR fixed effect delay model

NPCR Fixed Effect Model - Prostate

Year of Diagnosis	Crude	Age Standardized	Delay Adjusted	Delay Adjusted Age Standardized
1998	134.63	152.14	134.63	152.14
1999	144.40	162.21	144.95	162.83
2000	145.86	162.64	146.82	163.71
2001	147.66	163.38	149.25	165.14
2002	148.04	161.41	150.34	163.91
2003	134.70	144.86	137.63	148.01
2004	132.08	140.04	135.92	144.12
2005	132.53	138.22	137.91	143.84
2006	144.08	147.81	152.35	156.29
2007	144.00	144.36	158.32	158.71

NPCR Fixed Effect Model - Breast

Year of Diagnosis	Crude	Age Standardized	Delay Adjusted	Delay Adjusted Age Standardized
1998	162.71	152.92	162.71	152.92
1999	165.69	154.78	166.12	155.19
2000	161.65	150.15	162.38	150.83
2001	160.46	148.07	161.48	149.01
2002	158.08	144.77	159.50	146.06
2003	152.80	139.03	154.63	140.69
2004	152.73	138.09	155.12	140.26
2005	149.97	134.49	153.26	137.44
2006	151.99	135.59	156.53	139.65
2007	152.35	134.85	160.05	141.67

NPCR Fixed Effect Model - Lung

Year of Diagnosis	Crude	Age Standardized	Delay Adjusted	Delay Adjusted Age Standardized
1998	71.32	70.78	71.32	70.78
1999	69.31	68.58	69.32	68.59
2000	68.01	67.08	68.05	67.11
2001	67.63	66.46	67.70	66.53
2002	67.29	65.62	67.41	65.74
2003	66.55	64.34	66.73	64.51
2004	65.77	63.12	66.07	63.41
2005	65.57	62.40	66.20	63.01
2006	64.08	60.52	65.33	61.70
2007	61.92	57.85	65.13	60.85

NPCR Fixed Effect Model - Colon & Rectum

Year of Diagnosis	Crude	Age Standardized	Delay Adjusted	Delay Adjusted Age Standardized
1998	60.08	59.92	60.08	59.92
1999	58.62	58.21	58.66	58.25
2000	57.11	56.42	57.20	56.50
2001	55.82	54.88	55.97	55.02
2002	54.41	53.07	54.63	53.28
2003	53.51	51.69	53.82	51.98
2004	51.94	49.78	52.37	50.19
2005	49.84	47.30	50.53	47.95
2006	47.91	45.13	48.98	46.14
2007	45.83	42.67	47.97	44.66

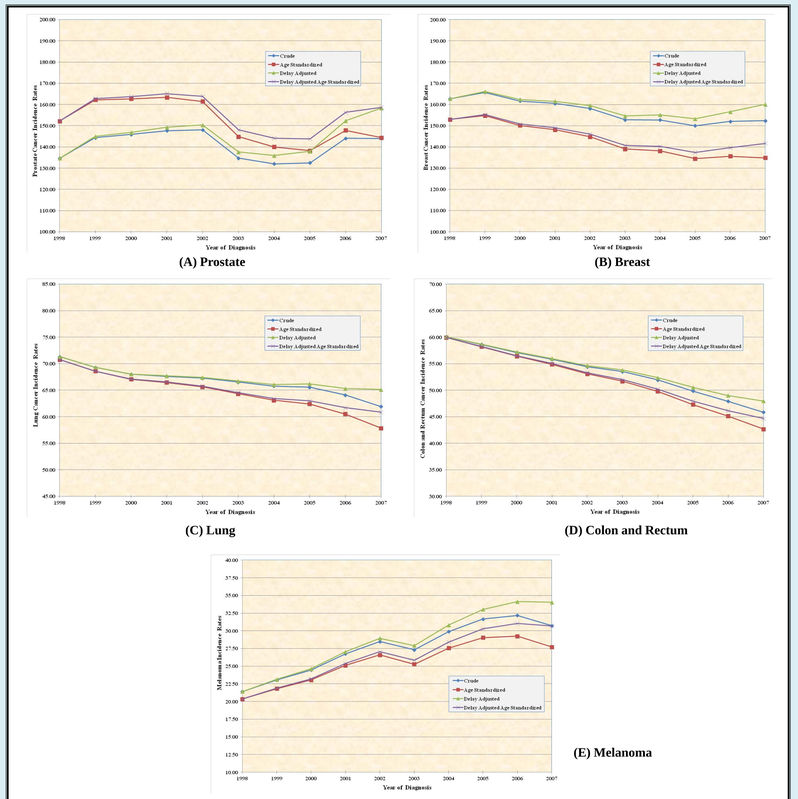

Figure 1: Estimated incidence rates from 5cancer sites using the NPCR fixed effect delay model. (A) Estimated prostate cancer incidence rates. (B) Estimated breast cancer incidence rates. (C) Estimated lung cancer incidence rates. (D) Estimated colon and rectum cancer incidence rates. (E) Estimated melanoma incidence rates.

Conclusions

1. The NPCR fixed effect delay model is extended from the NCI fixed effect delay model with simpler model specifications. The estimation process is stable and less prone to non-convergence.
2. The NPCR fixed effect delay model and the NCI fixed effect delay model agreed with each other in the predicted reporting delay distributions and the corresponding adjustments to the cancer incidence rates.
3. The NPCR fixed effect delay model is a good alternative approach to model reporting delay at the national level in the NPCR data.
4. The NPCR fixed effect delay model can be applied to states if the data quality at the state level meet the requirements for modeling.

References

1. Brookmeyer R, Liao J. The analysis of delays in disease reporting: methods and results for the acquired immunodeficiency syndrome. American Journal of Epidemiology 1990; 132:355–365.

2. Midthune DN, Fay MP, Clegg LX, and Feuer EJ. Modeling reporting delays and reporting corrections in cancer registry data. Journal of the American Statistical Association 2005; 100(469):61–70.

Acknowledgement

Special thanks to :
1. Douglas Midthune, Mathematical statistician at NCI
2. Dr. Ruiguang Song, Mathematical statistician at CDC/NCHHSTP/DHAP/QSDMB

NPCR NATIONAL PROGRAM OF CANCER REGISTRIES

ICF INTERNATIONAL

CDC