

Data Sources

For this study, we created a new linked research database for breast cancer cases diagnosed under age 65. Our linked database consists of 5 state cancer registries (California, Georgia, Kentucky, Ohio, and New York) containing clinical variables linked with claims data maintained by HealthCore Inc. (Wilmington, DE), an independent subsidiary to WellPoint, Inc, an independent licensee of the Blue Cross and Blue Shield Association in these 5 states plus 9 other states. We selected these 5 states (rather than all 14) because of their experience with prior linkage efforts, the quality of their registry data, inclusion of large numbers of WellPoint plan members, and cost efficiency. We then linked the Recurrence Score (RS) results for all tested cases in our linked sample through collaboration with Genomic Health Inc. (Redwood City, CA) who owns the patent for the Oncotype Dx test and conducts all testing in the US. All participating registries, HealthCore, and Georgetown obtained all necessary IRB and HIPAA approvals for this linkage.

The California, Georgia and Kentucky registries participate in the Surveillance, Epidemiology and End Results (SEER) program of the NCI. The New York State Cancer Registry (NYSCR) and the Ohio Cancer Incidence Surveillance System (OCISS) are part of the CDC funded National Program of Cancer Registries (NPCR). The NYSCR, one of the largest central cancer registries in the nation, has attained North American Association of Central Cancer Registries (NAACCR) gold certification for diagnosis years 1998-2011. The Ohio registry achieved Silver certification for those years. All 5 registries have participated in multiple data linkages with state, federal and researcher databases.

The HealthCore Data Environment includes the HealthCore Integrated Research Database (HIRDSM). The HIRD is a longitudinal medical and pharmacy claims, enrollment, and outpatient laboratory results database that includes data from commercially insured members' healthcare encounters and is owned and operated by HealthCore. The HIRD contains fully adjudicated paid claims processed for all non-capitated outpatient, emergency department, and inpatient encounters (including claims for laboratory tests), along with outpatient pharmacy claims for members with eligibility at the time of service. The data in the HIRD is fully identifiable which allows HealthCore, with appropriate waivers of HIPAA Authorization and/or Institutional

Review Board approval, to abstract inpatient and outpatient medical records, identify and contact providers and members for survey research, collect prospective data, and link data to registries and national vital records.

Matching Procedures (Will be online appendix only)

The starting study population was identified using claims data from the HealthCore Integrated Research Database (HIRDSM). We selected all women residing in the states of New York, California, Georgia, Kentucky, and Ohio who had at least two claims for breast cancer diagnosis (ICD-9 code 174.xx) at least 90 days apart between January 1, 2006 and April 30, 2012, and were between ages 24 and 64 on the first claim date with the diagnosis.

This study population (for each state) was sent to each of the 5 state cancer registries for matching. Except for Ohio, in which the linkage was done by Kentucky, all other states performed the linkage onsite. Identifiers such as social security number (SSN), names, date of birth, and street address at time of cancer diagnosis were used by state cancer registries to match these cases against all potential incident cases recorded in their statewide registry database through a probabilistic approach. Upon completion of linkages, clinical and diagnostic information for the matched cases were sent to HealthCore and then reviewed to ensure completeness and quality.

Table A: Linkage Results of HealthCore (HC) Data with Cancer Registry Data by State

State	HC data	Matched	% Matched
California	16,849	7,290	43.3
Georgia	5,829	2,224	38.2
New York	8,262	2,497	30.2
Kentucky	3,229	1,203	37.3
Ohio	6,913	2,850	41.2
Overall	41,082	16,064	39.1

The results of the linkage are presented in Table A. Overall, 39.1% of cancer cases identified using the HealthCore claims data were successfully matched with cases in the state cancer registries. California (43.3%)

and Ohio (41.2%) have higher match rates, followed by Georgia (38.2%) and Kentucky (37.3%), and New York (30.2%) has the lowest match rate. The primary reason for the observed linkage rate was the high frequency of “prevalent” cases, that is, most women identified by HealthCore had been diagnosed with breast cancer prior to 2006.

Genomic Health, Inc. also sent identifying information (names, date of birth) for all women residing in the 5 states who were recipients of the Oncotype Dx test for whom test results were available. HealthCore used the personal identifiers provided by Genomic Health to conduct a deterministic linkage. Upon completion of the linkage, information on Oncotype test results and other clinical data elements from Genomic Health were provided to HealthCore for incorporation into the final linked data set that combined registry information, claims data, and Oncotype Dx test results. Among 16,064 linked cases between HealthCore and cancer registries, 2,527 (15.7%) cases were successfully matched to women tested by Genomic Health. Some of these linked cases were not eligible for this analysis or had multiple tests. When women had multiple tests within a 90-day period of initial diagnosis, we selected the highest RS.

Appendix II. Sensitivity analyses for HER2 status, nodal involvement, and eligible patients

	<u>w/o known HER2 (n=4980)</u>			<u>only node negative (n=7054)</u>			<u>only N1 and N1Mic (n=2390)</u>			<u>Eligible (n=6546)</u>		
	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P	OR	95% CI	P
Year Diagnosed			<.0001			<.0001			<.0001			<.0001
<i>2006-2007(ref)</i>												
<i>2008-2009</i>	1.56	1.29 1.90		1.61	1.41 1.85		2.39	1.54 3.71		1.66	1.45 1.91	
<i>2010-2012</i>	1.67	1.37 2.04		1.35	1.17 1.57		4.07	2.64 6.28		1.41	1.21 1.64	
Age at diagnosis			<.0001			<.0001			0.2859			<.0001
<i>24-39</i>							1.12	0.54 2.33				
<i>40-49</i>							1.44	0.91 2.27				
<i>50-59</i>							1.48	0.96 2.26				
<i>60-64(ref)</i>												
Race - Ethnicity			0.0916			0.0231			0.0112			0.0488
<i>NH White(ref)</i>												
<i>NH Black</i>	0.62	0.41 0.96		0.66	0.48 0.90		0.36	0.12 1.03		0.69	0.50 0.94	
<i>Asian/Pac. Islanders</i>	0.86	0.66 1.12		0.84	0.66 1.07		0.61	0.32 1.15		0.82	0.64 1.05	
<i>Hispanic</i>	0.84	0.61 1.17		0.85	0.63 1.15		0.30	0.12 0.80		0.88	0.65 1.20	
Marital Status			0.914			0.7426			0.1989			0.9061
<i>Not Married</i>	0.99	0.85 1.15		1.02	0.90 1.15		1.24	0.89 1.72		1.01	0.89 1.14	
<i>Married(ref)</i>												
Prior Cancer			0.1373			0.2544			0.3292			0.197
<i>No(ref)</i>												
<i>Yes</i>	0.76	0.53 1.09		0.87	0.68 1.11		1.45	0.69 3.03		0.85	0.67 1.09	
State			0.0002			<.0001			0.2333			<.0001
<i>CA</i>	0.63	0.47 0.83		0.57	0.47 0.70		1.32	0.79 2.19		0.56	0.46 0.69	
<i>GA(ref)</i>												
<i>KY</i>	0.82	0.58 1.17		0.76	0.59 0.97		0.97	0.50 1.88		0.76	0.59 0.97	
<i>NY</i>	1.05	0.75 1.46		0.82	0.67 1.00		0.92	0.52 1.63		0.80	0.65 0.99	
<i>OH</i>	0.94	0.68 1.32		0.94	0.77 1.14		0.71	0.41 1.23		0.96	0.79 1.17	
Area			0.5654			0.024			0.257			0.0313
<i>Rural</i>	0.90	0.62 1.30		0.73	0.55 0.96		0.66	0.32 1.36		0.74	0.56 0.97	

<i>Urban(ref)</i>																
Median Household Income (area, Quintiles, Low to High)				0.4019				0.4023				0.6548				0.4564
<i>1 lowest (ref)</i>																
2	0.97	0.75	1.26		1.12	0.93	1.34		0.86	0.52	1.43		1.10	0.91	1.33	
3	1.16	0.90	1.48		1.15	0.96	1.39		1.07	0.66	1.74		1.13	0.94	1.37	
4	1.17	0.91	1.51		1.19	0.98	1.44		0.86	0.51	1.46		1.18	0.97	1.43	
<i>5 highest</i>	1.14	0.89	1.48		1.20	0.99	1.47		0.75	0.44	1.29		1.20	0.98	1.47	
Out-of-Pocket Pharmacy Costs (Quintiles, Low to High)				<.0001				<.0001				0.0002				<.0001
<i>1 lowest (ref)</i>																
2	0.99	0.79	1.25		1.08	0.90	1.31		0.98	0.58	1.65		1.10	0.91	1.34	
3	1.14	0.91	1.43		1.17	0.97	1.41		1.75	1.06	2.90		1.18	0.98	1.43	
4	1.45	1.16	1.82		1.51	1.26	1.81		1.68	1.01	2.79		1.54	1.28	1.85	
<i>5 highest</i>	1.66	1.33	2.07		1.58	1.31	1.90		2.68	1.62	4.43		1.64	1.36	1.98	
Stage at Diagnosis				0.0924				0.1851								0.2268
<i>Stage I (ref)</i>									*****	*****	*****	*****				
<i>Stage II</i>									*****	*****	*****	*****				
Nodal Involvement				<.0001								<.0001				
<i>N0 (ref)</i>					*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****	*****
N1mi	0.69	0.50	0.96		*****	*****	*****	*****	3.33	2.46	4.50		*****	*****	*****	*****
N1	0.23	0.17	0.30		*****	*****	*****	*****	Ref				*****	*****	*****	*****
HER2 Status				<.0001				<.0001				0.0005				<.0001
<i>Positive</i>	0.29	0.23	0.39		0.32	0.24	0.42		0.14	0.04	0.46		*****	*****	*****	*****
<i>Negative(ref)</i>																
<i>Borderline/unknown</i>	0.88	0.56	1.38		0.49	0.42	0.57		0.62	0.42	0.91		0.49	0.42	0.57	
Hormone Receptor Status				0.0013				<.0001				0.079				<.0001
<i>ER and PR both positive(ref)</i>																
<i>At least one positive</i>	0.70	0.57	0.87		0.69	0.58	0.82		0.63	0.38	1.05		0.70	0.59	0.84	
Histological Grade				0.0499				0.1342				<.0001				0.3105

<i>1-2: Well/moderately differentiated (ref)</i>																
<i>3: Poorly or not differentiated</i>		0.83	0.69	1.00		0.90	0.77	1.04		0.40	0.26	0.60		0.93	0.80	1.08
One year Comorbidities					0.032				0.0014				0.1575			0.0037
<i>0 (ref)</i>																
<i>1 or more</i>		1.28	1.02	1.59		1.34	1.12	1.60		1.44	0.87	2.37		1.31	1.09	1.58
Age at diagnosis	Stage at Diagnosis				0.0004				0.0006							0.0017
24-39	<i>Stage I</i>	2.09	1.13	3.86		1.82	1.06	3.12		*****	*****	*****	*****	2.00	1.14	3.51
	<i>Stage II</i>									*****	*****	*****	*****			
40-49	<i>Stage I</i>	1.65	1.25	2.18		1.81	1.41	2.33		*****	*****	*****	*****	1.73	1.34	2.23
	<i>Stage II</i>									*****	*****	*****	*****			
50-59	<i>Stage I</i>	1.08	0.84	1.40		1.27	1.01	1.59		*****	*****	*****	*****	1.28	1.01	1.61
	<i>Stage II</i>									*****	*****	*****	*****			
60-64	<i>Stage I</i>	0.75	0.54	1.05		0.81	0.60	1.10		*****	*****	*****	*****	0.82	0.60	1.13
	<i>Stage II</i>									*****	*****	*****	*****			

***** N/A for that cohort