Spatial Analysis in Cancer Surveillance: Issues in Identifying Geographic Targets for Screening Interventions

Recinda L Sherman[1], Kevin A Henry[2], Stacey L Tannenbaum[1], Daniel J Feaster[1], Erin Kobetz[1], David J Lee[1]

1 University of Miami Miller School of Medicine, 2 University of Utah, Department of Geography

I OVERVIEW:

Colorectal cancer (CRC) is one of the most common invasive cancers in the US. Routine screening reduces mortality and some screening modalities have the potential to reduce incidence, making CRC potentially preventable. Despite strong evidence that CRC screening saves lives, screening rates remain low, particularly among minorities. This translates to an increased burden of CRC diagnosed at a late-stage for these groups.

Geographically targeting high risk populations for enhanced CRC screening efforts is pragmatic public health policy. While population-based screening interventions are designed to be innocuous for low risk populations, public health resources are limited. There will be geographic disparities in cancer rates, but the challenge is to correctly identify which areas have potential for reduction of excess disease burden through public health intervention.

CRC was analyzed with SaTScan to identify clusters of late-stage CRC in Florida using several methods and scales. **Results indicated much of the analysis was underpowered and there is no one scale or method that will identify all clusters of statistical or public health significance.**

II METHODS:

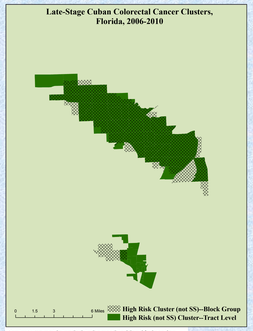
This was a Florida, population-based, ecologic study on the geographic distribution of CRC. CRC Cases diagnosed 1996-2010, age 50+, adenocarcinomas, and geocoded to 2010 census tract or block group based on street address were used. "Early" was defined as *in situ* or localized at diagnosis, and, because unknown stage has poor prognosis, regional, distant, and unknown were "late". US Census 2010 population data was used.

Spatial scans in SaTScan ver 9.1.1 were used to detect clusters of late-stage at diagnosis CRC at both the census tract and block group-level. Circular and elliptic shaped scan windows were used simultaneously and analyses were conducting at multiple scales (1%, 2%, and 5-50% of the population at risk in 5% increments as maximum potential cluster size). Due to extreme computational requirements, Gumbel based p-values were used which resulted in comparable p-values in less time.

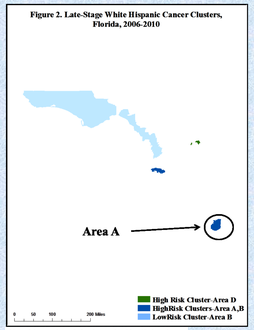
Two spatial scan probability models were used: Poisson Model (PM—identifies areas of late-stage incidence using age-adjusted rates) and Bernoulli Model (BM—identifies areas of late-stage incidence using a ratio of late:early counts). BM was used to identify high risk clusters for late-stage at diagnosis (compared to early-stage) CRC for Blacks, White Hispanics, White Non-Hispanics and Cubans. PM was used to identify high and low risk clusters for late-stage at diagnosis CRC for the same race/ethnicity groups, except Cubans due to lack of census data. The most like cluster functionality was used, p=.05, maximum iterations 15.

Table 1. Colorectal Cluster Summary, Floridians, 2006-2010

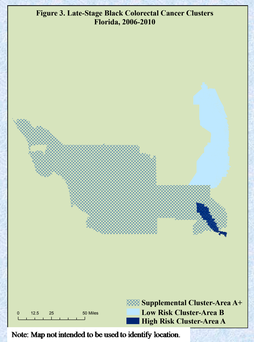
% Pop (Maximum Cluster Size)	Mapped Location	RR (Cluster) Block Group	RR (Cluster) Tract	RR Range (Local) Block Group	RR Range (Local) Tract	P value Block Group	P value Tract
Cuban Cluster Summary (Bernoulli Only)							
1		1.61	1.61	0, 1.6; 0.72	0, 1.6; 0.76	0.40	0.95
2		1.61	1.62	0, 1.6; 0.56	0, 1.6; 0.76	0.61	0.48
5		1.61	1.62	0, 1.6; 0.56	0, 1.6; 0.76	0.68	0.53
10		1.61	1.62	0, 1.6; 0.56	0, 1.6; 0.76	0.70	0.56
15		1.61	1.62	0, 1.6; 0.56	0, 1.6; 0.76	0.70	0.56
20		1.61	1.62	0, 1.6; 0.56	0, 1.6; 0.76	0.70	0.57
25	Area A	1.61	1.62	0, 1.6; 0.56	0, 1.6; 0.76	0.71	0.57
30		1.61	1.62	0, 1.6; 0.56	0, 1.6; 0.76	0.71	0.57
35		1.61	1.62	0, 1.6; 0.56	0, 1.6; 0.76	0.71	0.57
40		1.61	1.62	0, 1.6; 0.56	0, 1.6; 0.76	0.71	0.58
45		1.61	1.62	0, 1.6; 0.56	0, 1.6; 0.76	0.70	0.58
50		1.61	1.62	0, 1.6; 0.56	0, 1.6; 0.76	0.90	0.58
White Hispanic Cluster Summary (Bernoulli)							
1	Area A	1.54	1.52	0, 1.7; 0.8	0, 1.7; 0.7	0.45	0.23
2	Area A	1.54	1.52	0, 1.7; 0.8	0, 1.7; 0.7	0.48	0.27
5	Area A	1.54	1.52	0, 1.7; 0.8	0, 1.7; 0.7	0.50	0.29
10	Area A	1.54	1.52	0, 1.7; 0.8	0, 1.7; 0.7	0.51	0.30
15	Area A	1.54	1.52	0, 1.7; 0.8	0, 1.7; 0.7	0.52	0.30
20	Area A	1.54	1.52	0, 1.7; 0.8	0, 1.7; 0.6	0.52	0.30
25	Area A	1.54	1.35	0, 1.7; 0.8	0, 10.2; 1.5	0.52	0.00
25	Area D	--	1.25	--	0, 15.3; 2.0	--	0.12
30	Area A	1.54	1.52	0, 1.7; 0.8	0, 1.7; 0.7	0.52	0.31
35	Area A	1.54	1.52	0, 1.7; 0.8	0, 1.7; 0.7	0.52	0.31
40	Area A	1.54	1.52	0, 1.7; 0.8	0, 1.7; 0.7	0.52	0.31
45	Area A	1.54	1.52	0, 1.7; 0.8	0, 1.7; 0.7	0.52	0.31
50	Area A	1.54	1.52	0, 1.7; 0.8	0, 1.7; 0.7	0.50	0.31
White Hispanic Cluster Summer (Poisson)							
1					0, 233.8;		
1	Area B	0.24	0.25	0, 144.6; 6.4	13.6	0.04	0.37
					0, 236.8;		
2	Area B	0.32	0.25	0, 144.6; 5.8	13.8	0.08	0.49
2	Area A	--	2.00	--	0, 10.2; 2.5	--	0.04
5	Area A	1.56	1.70	0, 27.7; 3.5	0, 10.2; 2.1	0.01	0.04
5	Area A	1.55	--	0, 52.8; 6.4	--	0.09	--
5	Area C	--	0.61	--	0, 22.5; 1.9	--	0.26
10	Area A	1.46	1.44	0, 27.7; 3.22	0, 10.2; 1.7	0.00	0.00
10	Area A	--	1.38	--	0, 6.8; 1.1	--	0.06
10	Area A	--	1.38	--	0, 5.0; 0.8	--	0.22
10	Area C	0.62	--	0, 15.7; 1.8	--	0.06	--
15	Area A	1.40	1.35	0, 51.4; 3.6	0, 10.2; 1.5	0.00	0.00
15	Area A	1.41	1.39	0, 45.1; 3.5	0, 6.9; 1.1	0.00	0.00
15	Area A	--	1.58	--	0, 8.6; 1.5	--	0.04
15	Area C	1.70	1.38	0, 22.0; 3.0	0, 15.6; 2.1	0.07	0.09
15	Area D	--	6.46	--	0, 19.9; 6.9	--	0.93
20	Area A	1.30	1.38	0, 51.4; 4.4	0, 10.2; 1.5	0.00	0.00
20	Area A	1.29	1.36	0, 16.4; 1.9	0, 5; 0.8	0.06	0.06
20	Area A	--	1.45	--	0, 9.4; 1.5	--	0.17
20	Area C	--	1.60	--	0, 15.8; 2.1	--	0.02
25	Area A	1.39	1.35	0, 51.4; 3.7	0, 10.2; 1.5	0.00	0.00
25	Area C	--	1.54	--	0, 15.3; 2.0	--	0.12
30	Area A	1.53	1.34	0, 5.1; 1.1	0, 10.2; 1.3	0.11	0.00
30	Area A	--	1.53	--	0, 14.4; 2.2	--	0.02
30	Area C	--	1.57	--	0, 15.5; 2.1	--	0.06
35	Area A	1.38	1.35	0, 42.7; 2.8	0, 10.2; 1.3	0.00	0.00
35	Area A	1.53	1.53	0, 60.1; 1.8	0, 8.7; 1.6	0.20	0.41
35	Area C	1.62	1.60	0, 47.5; 4.0	0, 15.7; 2.1	0.03	0.03
40	Area A	1.39	1.36	0, 42.7; 2.8	0, 10.2; 1.3	0.00	0.00
40	Area A	1.54	1.51	0, 0; 61.0; 5.5	0, 26.0; 2.9	0.18	0.22
40	Area C	1.62	1.63	0, 47.7; 4.0	0, 16.0; 2.1	0.02	0.01
45	Area A	1.39	1.36	0; 84.2; 3.9	0, 10.2; 1.3	0.00	0.00
45	Area A	--	1.51	--	0, 26.0; 3.0	--	0.22
45	Area C	1.61	1.63	0, 49.1; 4.1	0, 15.9; 2.1	0.00	0.01
50	Area A	1.39	1.36	0, 84.2; 3.9	0, 106.3; 4.8	0.00	0.00
50	Area C	1.69	1.71	0, 49.4; 4.1	0, 16.5; 2.2	0.00	0.00
50	Area D	--	6.72	--	0, 20.6; 7.2	--	0.85
Black Cluster Summary (Bernoulli)							
1	Area A	--	1.60	--	0, 1.7; 0.8	--	0.18
2	Area A	1.55	1.48	0, 1.7; 0.6	0, 1.7; 0.8	0.10	0.06
5	Area A	1.37	1.43	0, 1.7; 0.7	0, 1.7; 0.8	0.04	0.05
10	Area A	1.37	1.43	0, 1.7; 0.7	0, 1.7; 0.8	0.04	0.05
15	Area A	1.37	1.43	0, 1.7; 0.7	0, 1.7; 0.8	0.04	0.05
20	Area A	1.37	--	0, 1.7; 0.7	--	0.04	0.05
20	Area A+	--	1.19	--	0, 1.7; 0.7	--	--
25	Area A	1.37	--	0, 1.7; 0.7	--	0.04	0.05
25	Area A+	--	1.19	--	0, 1.7; 0.7	--	--
30	Area A	1.37	--	0, 1.7; 0.7	--	0.04	0.05
30	Area A+	--	1.19	--	0, 1.7; 0.7	--	--
35	Area A	1.37	--	0, 1.7; 0.7	--	0.04	0.05
35	Area A+	--	1.19	--	0, 1.7; 0.7	--	--
40	Area A	1.37	--	0, 1.7; 0.7	--	0.04	0.05
40	Area A+	--	1.19	--	0, 1.7; 0.7	--	--
45	Area A	1.37	--	0, 1.7; 0.7	--	0.04	0.05
45	Area A+	--	1.19	--	0, 1.7; 0.7	--	--
50	Area A	1.38	--	0, 1.7; 0.7	--	0.04	--
50	Area A+	--	1.19	--	0, 1.7; 0.7	--	0.05
Black Cluster Summary (Poisson)							
1	Area A	--	4.00	--	1.7, 6.4; 1.7	--	0.03
2	Area B	0.35	0.00	0, 32.4; 2.0	0, 0; 0;	0.32	0.33
2	Area A	--	4.00	--	1.7, 6.4; 1.7	--	0.03
5	Area A	1.55	1.43	0, 45.3; 4.1	0, 1.7; 0.8	0.27	0.05
5	Area B	0.38	--	0, 32.4; 2.1	--	0.00	--
10	Area A	1.55	1.53	0, 45.3; 4.6	0,33.3; 4.2	0.12	0.03
10	Area B	0.38	0.42	0, 32.4; 2.1	0,11.2; 1.2	0.00	0.00
15	Area A	1.51	1.53	0, 45.3; 4.6	0,33.3; 4.2	0.12	0.03
15	Area B	0.38	0.42	0, 32.4; 2.1	0,11.2; 1.2	0.00	0.00
20	Area A	1.55	1.53	0, 45.3; 4.6	0,33.3; 4.2	0.12	0.03
20	Area B	0.38	0.42	0, 32.4; 2.1	0,11.2; 1.2	0.00	0.00
25	Area A	1.55	1.53	0, 45.3; 4.6	0,33.3; 4.2	0.12	0.03
25	Area B	0.38	0.42	0, 32.4; 2.1	0,11.2; 1.2	0.00	0.00
25	Area A	1.55	1.53	0, 45.3; 4.6	0,33.3; 4.2	0.12	0.03
30	Area A	1.55	1.53	0, 45.3; 4.6	0,33.3; 4.2	0.12	0.03
35	Area A	1.55	1.53	0, 45.3; 4.6	0,33.3; 4.2	0.12	0.03
35	Area B	0.38	0.42	0, 32.4; 2.1	0,11.2; 1.2	0.00	0.00
40	Area A	1.51	1.53	0, 45.3; 4.6	0,33.3; 4.2	0.12	0.03
40	Area B	0.38	0.42	0, 32.4; 2.1	0,11.2; 1.2	0.00	0.00
45	Area A	1.51	1.53	0, 45.3; 4.6	0,33.3; 4.2	0.12	0.03
45	Area B	0.38	0.42	0, 32.4; 2.1	0,11.2; 1.2	0.00	0.00
50	Area A	--	1.53	--	0,33.3; 4.2	--	0.03
50	Area B	0.38	0.42	0, 32.4; 2.1	0,11.2; 1.2	0.00	0.01



Note: Map not intended to be used to identify location.



Note: Map not intended to be used to identify location.



Note: Map not intended to be used to identify location.

III RESULTS:

Cuban analysis did not identify statistically significant clusters. However, with exception of the smallest scale, all methods identified the same area in S Florida as the most likely cluster (**Figure 1**). The relative risk (RR), 1.6, for the cluster was not statistically significant but was consistent at all scales (block group versus tract and % of population as maximum cluster size) as was the location with the exception of smallest scale (1% of population at risk) (**Table 1**). The overlap was exact for all analysis done at the same census boundary-level, but the block group-level analysis cluster included supplemental outlying blocks (**Figure 1**). The location overlap resulted in comparable local risk variation as measured by range of local risk and standard deviation. In general, the p-values were larger for the analysis done at block group-level and larger scale (greater maximum cluster size).

White Hispanic analysis detected three statistically significant clusters of high risk of late-stage CRC at p=.05, 2 high risk and 1 low risk (**Table 1**). All BM analysis consistently identified the same area of high risk clustering but only the analysis at the scale 25% of the population was statistically significant. In general, p-values for the block group-level analysis were smaller than for the tract-level and the lowest scale analysis resulted in higher p-values. Ringed clusters were identified in high risk Area A, although the RR were statistically different for these clusters, the difference between a RR 1.5 and 1.4 does not have practical significance for ecologic studies. All island clusters with geographic overlap and adjacency are identified as the cluster Area A (**Figure 1**). This area overlaps the Black and Cuban clusters. A 4th area, non-statistically significant Area D, was identified at the 25% of population scale in BM and 15% and 50% scale in PM for tract-level analysis.

Black analysis detected two statistically significant late-stage clusters at p=.05, one high risk area and one low risk. These clusters were also in S Florida, and the high risk area was in close proximity/overlapped the non-significant area identified in the Cuban analysis. BM analysis at tract-level detected a larger geographic area at high risk of late-stage colorectal diagnosis (**Figure 2**). The low risk cluster, identified using PM, did not overlap with any high risk area identified using BM. RR for each cluster, approximately 1.5 and 0.4, were similar among all methods, although the BM RR range was lower, 1.2-1.4. P-values were similar for both tract and block group analysis for the BM, with the values generally lower at the smaller scales. The PM block group-level analysis did not result in a statistically significant clusters, but the p-values for the low risk cluster were similar regardless of census boundary-level. In general, the local variation in risk was smaller at the tract-level.

IV CONCLUSIONS:

Incorporating cluster detection into existing disease surveillance activities can proactively identify areas of high risk to target interventions and drive etiologic research. But there remains a need to establish best practices to ensure identified areas of concern are "real", that is, have the potential to be amenable through public health intervention or to contribute to etiologic knowledge. **Protocol should be established so that the "art" of analysis is replicable and reduces the potential for false positives.** The selected geographic areas must have real potential for attenuating excess CRC burden through increased screening efforts. Reliance on SaTScan parameter defaults, pre-determined cut-points, or cookie cutter analysis is not appropriate.