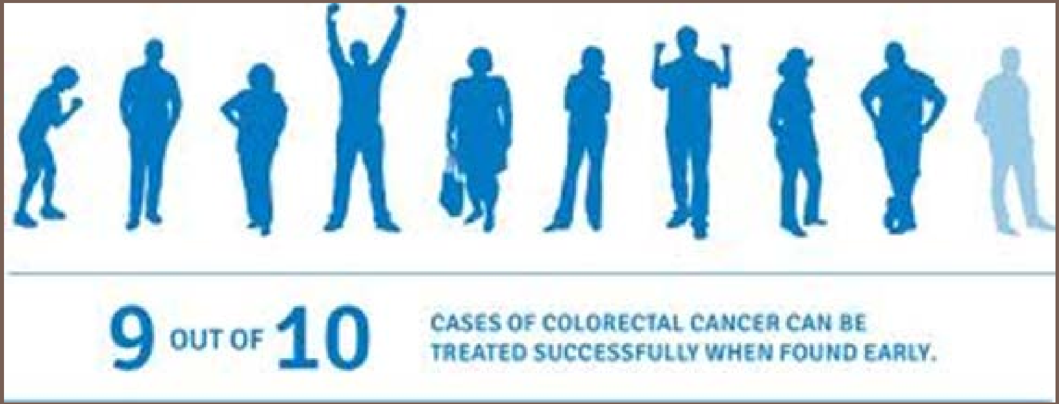
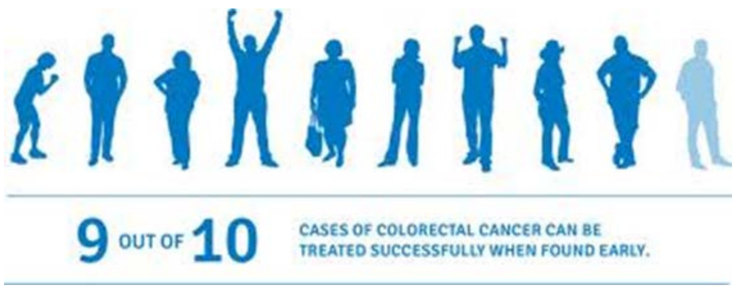


THE IMPACT OF MISSING STAGE AT DIAGNOSIS ON RESULTS OF GEOGRAPHIC RISK OF LATE-STAGE COLORECTAL CANCER

Recinda Sherman, MPH, PhD, CTR
NAACCR Annual Meeting, Wed June 25, 2014

Outline

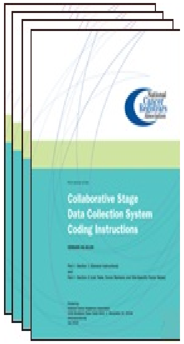
- Background: Missing data, Colorectal Cancer, Colorectal Cancer Screening
- Methods: Cluster Detection, Area-based measures, Distance analysis
- Results: Cluster Detection, Area-based measures, Distance analysis
- Conclusions
  - Future directions



Cancer Registry Purpose

"These data are then used to inform a wide variety of public health decisions and provide rich information for cancer diagnosis and treatment education."

- Guides policy and treatment decisions
  - Reduce the burden of cancer
- Error
  - Inappropriate public health response
  - Fail to protect population; waste of public funds
- Cause
  - Flawed research design, inappropriate assumptions, bias
  - **Data quality in registry data**



Missing data

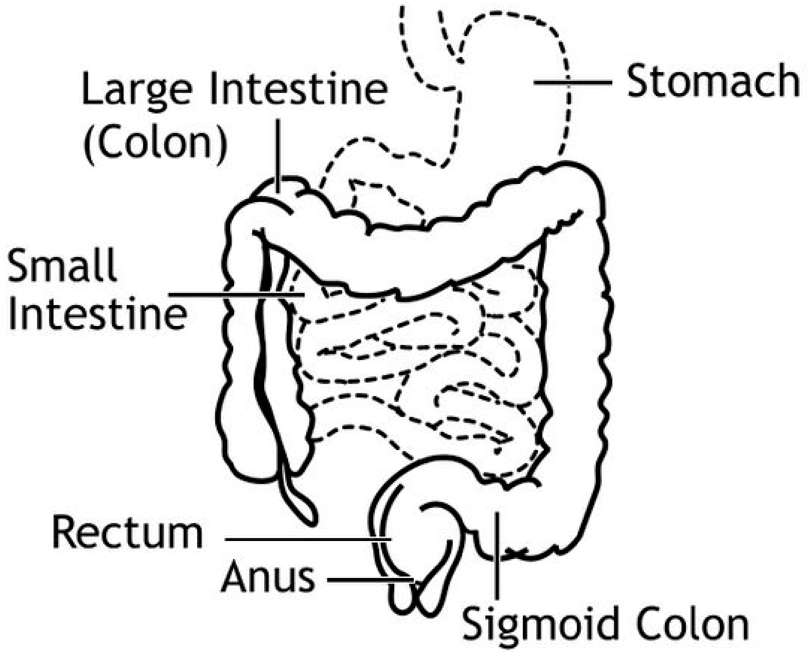
- Random – hard to correct during collection but established methods for analysis
  - Remove cases with missing data, random allocation of missing information, interpolation based on known data
- Systematic – "easy" to correct during collection but no established method for analysis
  - Cartographic selection bias
- Leads to reduced power and possibly biased results

Background: Cancer surveillance

- System of central (state) cancer registries
- Collect incidence, type, anatomic location, extent of disease, treatment, and outcomes
  - National standards, “Gold Standard”
- Systematic analysis of cancer data
  - Identify burden and trends of cancer
  - Generate hypotheses about cancer risk and etiology
- Guides policy and treatment decisions
  - Reduce the burden of cancer

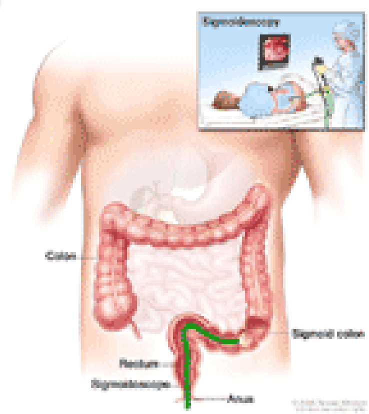
Background: Colorectal cancer

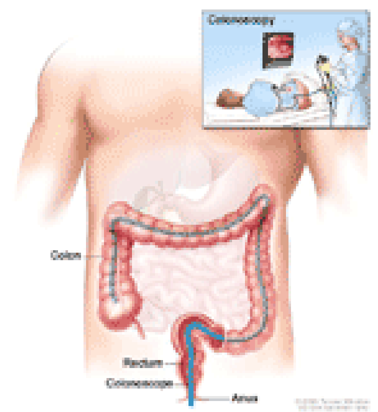
- 2nd Mortality
- 4th Invasive
- Cause ?
  - Multifactorial
  - Diet, exercise, HPV
- 95% adenocarcinomas
  - Pre-cancerous polyps



Background: Screening

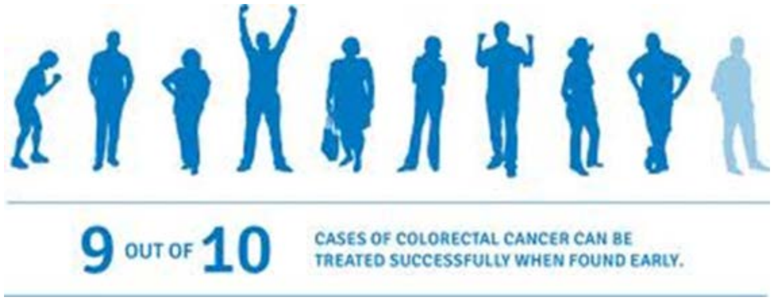
- Multimodal
  - FOBT, sigmoidoscopy, colonoscopy
  - 50+
- USPSTF Grade A recommendation
  - Primary and secondary prevention
- 50% general population
  - Blacks, lower
  - Hispanics, lowest





Stage of disease

- Common outcome measure: early vs late
  - Proxy for screening uptake
  - Proxy for prognosis
- Often missing/unknown
  - Lack of clinical assessment
    - Lack of connection to health services
    - Contraindicated (age, comorbidities)
    - Low survival, likely unscreened
  - Lack of collection
    - Unknown survival/screening



Aim of study

- Evaluate the impact of missing stage at diagnosis for colorectal cancer geospatial research
  - By race/ethnicity
- Three different methods of handling missing stage
  - Remove cases (most common approach)
  - Allocate based on demographics
  - Code all as late
- Three different cancer control questions
  - Where should we target a screening intervention?
  - How should we tailor the intervention based on demographics?

  Are disparities being driving by unequal proximity to clinical care?

Methods: Florida Cancer Data System

- NPCR, Incidence based, 1981+
- 2nd largest cancer registry in US
- 115,000 incident cancer cases annually
  - 200,000 reports, 150,000 Death, 1,000,000 Discharge
- Gold Certification
- System
  - Facilities, physicians, vital statistics, hospital discharge, medical billing, pathology
- Geocoding
  - 95% street level, 3% PO Box
  - Proprietary vendor, no manual follow-up

Methods: Case selection

- Colorectal cases
  - diagnosed 1996-2010
  - 1st primary, age 50+, adenocarcinomas
- Geocoded to 2010 tract or block group
  - based on street address
- Stage at diagnosis
  - Dichotomous “early” & “late”

Methods: Cluster detection

- **First:** Detect high risk clusters of late-stage at diagnosis CRC
  - Blacks, Cubans, White Hispanics, White Non-Hispanics
- SaTScan
  - Spatial Scan
  - Poisson Model
    - Late-stage incidence using age-adjusted rates
  - Bernoulli Model
    - ratio of counts late:early

- Details in upcoming *Preventing Chronic Disease* method paper: *Issues in applying spatial analysis tools in public health: an example of using SaTScan to identify geographic targets for colorectal cancer screening interventions*

Methods: Area-based measures

- 2006-2010 ACS aggregated dataset
  - Tract-level measure
  - Socio-demographic characteristics
- 2010 Florida BRFSS
  - County-level measure
  - Screening uptake
- Hierarchical, logistic regression models
  - Proc glimmix

Methods: Distance analysis

- Travel distance between patient's residence and facility that reported the case
  - Median and mean
- NAACCR Shortest Path Finder Tool
  - Road networks, TeleAtlas

Results:

	Black (n=3,779)			Hispanic White (n=4,989)			Non-Hispanic White (n=28,796)		
	Unknown			Unknown			Unknown		
	=Late	=Exclude	=Allocate	=Late	=Exclude	=Allocate	=Late	=Exclude	=Allocate
Total late cases	2,242	2,058	2,165	3,003	2,778	2,907	16,160	14,782	15,512
All combined	59%	57%	57%	60%	58%	58%	56%	54%	54%
Sex									
female	58%	56%	56%	60%	58%	58%	56%	54%	54%
male	61%	59%	59%	60%	58%	58%	56%	54%	54%
Age									
50-54	61%	60%	60%	62%	60%	60%	58%	57%	57%
55-59	64%	63%	63%	67%	65%	65%	60%	59%	59%
60-64	60%	58%	58%	62%	60%	60%	59%	58%	58%
65-69	57%	55%	55%	58%	57%	57%	56%	54%	54%
70-74	56%	54%	54%	61%	59%	60%	54%	52%	52%
75-79	59%	58%	58%	60%	58%	58%	54%	52%	51%
80-84	57%	53%	53%	57%	55%	55%	54%	51%	51%
85+	60%	55%	54%	58%	54%	54%	57%	53%	53%

Table 1. Distribution of late stage by method of handling unknowns, by race/ethnicity, sex, and age

Results: Cluster detection

	# clusters			# cases in cluster(s)			range of RR^			range of p		
	=Late	=Exclude	=Allocate	=Late	=Exclude	=Allocate	=Late	=Exclude	=Allocate	=Late	=Exclude	=Allocate
Black (B 30%)	1	1	1	581	58	423	1.2	1.5	1.2	0.05	0.1	0.4
Black (P 40%)	1	1	2	86	284	196	1.5	1.4	1.5- 4.2	0.03	0.1	<01 - 0.05
HW (B na)	na	na	na	na	na	na	na	na	na	na	na	na
HW (P50%)	2	3	2	1,860	1,555	1,455	1.4 - 1.7	1.4 - 1.7	1.4 - 1.6	<.001 - <.01	<.001 - 0.05	<.001 - 0.01
nonHW (B 15%)	3	2	4	4,692	3,704	3,380	1.1 - 1.2	1.1 - 1.1	1.1 - 1.2	<.001 - 0.03	<.001	<.001 - 0.02
nonHW (P 25%)	12	11	12	9,823	8,048	8,321	1.2 - 5.2	1.2 - 8.8	1.2-5.1	<.001 - 0.03	<.001 - 0.03	<.001 - 0.05

All = allocate B = Bernoulli, P = Poisson, * maximum cluster size (scale), ^RR = Relative Risk, HW= Hispanic White

note: none of the Bernoulli analysis was statistically significant for comparision.

Table 2. Late-stage at diagnosis colorectal cancer clusters, Florida 2006-2010

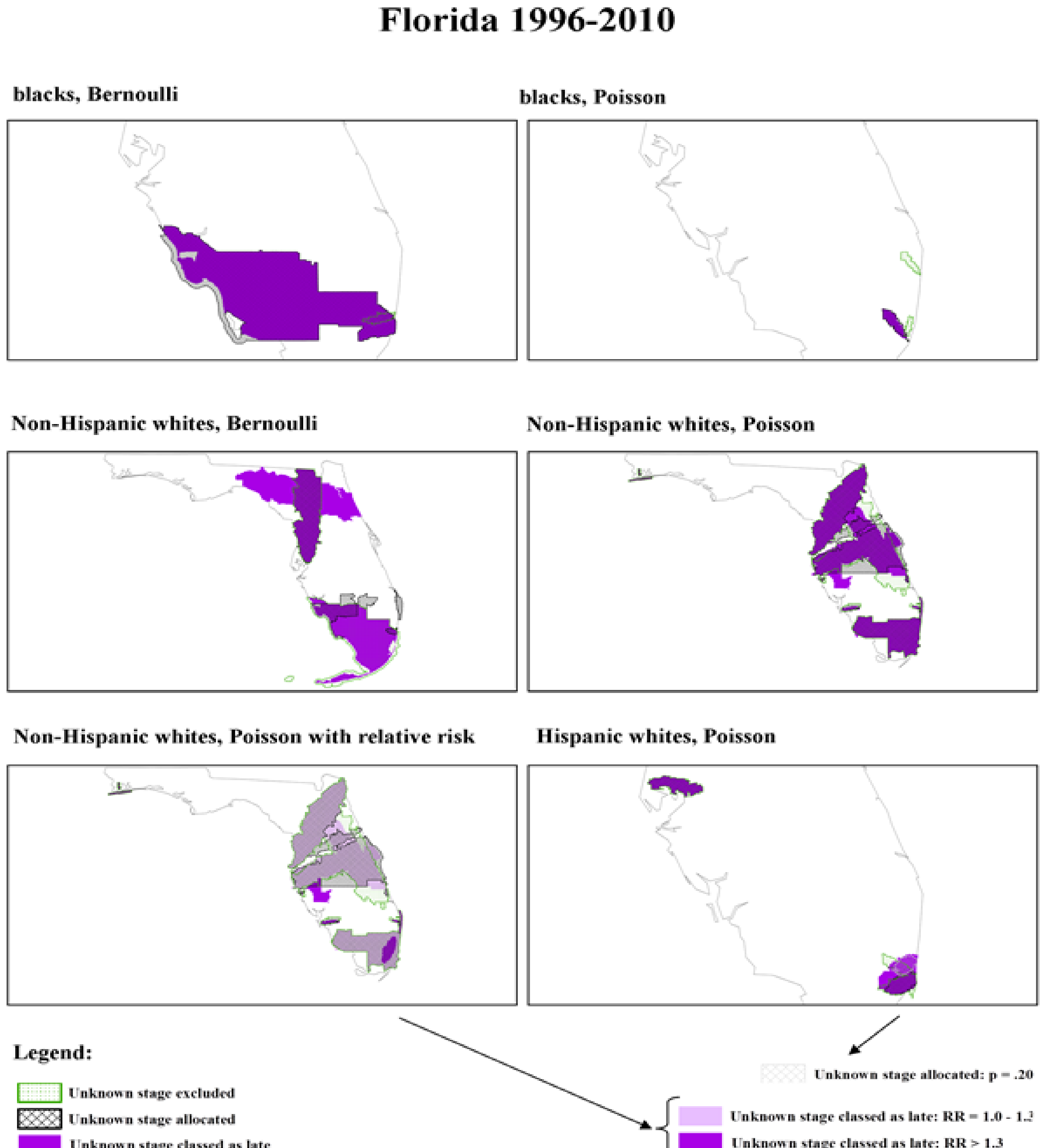


Figure 1. Comparison of tract-based cluster locations by method of handling unknowns, by race/ethnicity and Model type

Results: Area-based measures

Non-Hispanic White:Bernoulli Method	Unknown=Late			Unknown=Exclude			Unknown=Allocate		
Tract Level	OR	p-value	CI	OR	p-value	CI	OR	p-value	CI
% non-white	1.0	<.001	1.0, 1.0	1.0	<.001	1.0, 1.0	1.0	0.04	1.0, 1.0
% hispanic	1.0	0.03	1.0, 1.0	1.0	0.00	1.0, 1.0	*not included*		
% minority	*not included*			*not included*			*not included*		
% foreign born	1.1	<.001	1.1, 1.1	1.1	<.001	1.1, 1.1	1.0	<.001	1.0, 1.1
% not hs grad	*not included*			*not included*			*not included*		
% no English spoken	0.9	0.00	0.8, 1.0	0.9	0.00	0.8, 1.0	0.9	0.00	0.9, 1.0
% below poverty	1.0	0.00	1.0. 1.1	*not included*			1.0	0.05	1.0, 1.0
County Level									
% ever received sigmoidoscopy/colonscopy	0.8	<.001	0.8, 0.8	0.8	<.001	0.8, 0.8	1.0	0.0	0.9, 1.0
% received fobt last 2 years	*not included*			*not included*			1.1	<.001	1.1, 1.1
Non-Hispanic White: Poisson Method	**Unknown=Late**			**Unknown=Exclude**			**Unknown=Allocate**		
Tract Level	**OR**	**p-value**	**CI**	**OR**	**p-value**	**CI**	**OR**	**p-value**	**CI**
% non-white	*not included*			*not included*			*not included*		
% hispanic	*not included*			*not included*			*not included*		
% minority	1.1	<.001	1.1, 1.1	1.1	<.001	1.1, 1.1	1.1	<.001	1.0, 1.1
% foreign born	1.1	<.001	1.0, 1.1	*not included*			1.0	0.02	1.0, 1.1
% not hs grad	*not included*			1.0	0.01	1.0, 1.0	*not included*		
% no English spoken	0.8	<.001	0.8, 0.9	0.9	0.04	0.7, 0.8	0.8	<.001	0.8, 0.9
% below poverty	1.0	0.00	1.0, 1.0	1.0	0.03	1.0, 1.0	1.0	0.0	1.0, 1.0
County Level									
% ever received sigmoidoscopy/colonscopy	0.8	<.001	0.8, 0.8	0.8	<.001	0.8, 0.9	0.8	<.001	0.8, 0.8
% received fobt last 2 years	*not included*			1.0	0.03	0.9, 1.0	1.0	0.01	0.9, 0.9

Results: Distance analysis

Non-Hispanic White:Bernoulli Method	Unknown=Late			Unknown=Exclude			Unknown=Allocate		
Tract Level	OR	p-value	CI	OR	p-value	CI	OR	p-value	CI
% non-white	1.0	<.001	1.0, 1.0	1.0	<.001	1.0, 1.0	1.0	0.04	1.0, 1.0
% hispanic	1.0	0.03	1.0, 1.0	1.0	0.00	1.0, 1.0	*not included*		
% minority	*not included*			*not included*			*not included*		
% foreign born	1.1	<.001	1.1, 1.1	1.1	<.001	1.1, 1.1	1.0	<.001	1.0, 1.1
% not hs grad	*not included*			*not included*			*not included*		
% no English spoken	0.9	0.00	0.8, 1.0	0.9	0.00	0.8, 1.0	0.9	0.00	0.9, 1.0
% below poverty	1.0	0.00	1.0. 1.1	*not included*			1.0	0.05	1.0, 1.0
County Level									
% ever received sigmoidoscopy/colonscopy	0.8	<.001	0.8, 0.8	0.8	<.001	0.8, 0.8	1.0	0.0	0.9, 1.0
% received fobt last 2 years	*not included*			*not included*			1.1	<.001	1.1, 1.1

Non-Hispanic White: Poisson Method	Unknown=Late			Unknown=Exclude			Unknown=Allocate		
Tract Level	OR	p-value	CI	OR	p-value	CI	OR	p-value	CI
% non-white	*not included*			*not included*			*not included*		
% hispanic	*not included*			*not included*			*not included*		
% minority	1.1	<.001	1.1, 1.1	1.1	<.001	1.1, 1.1	1.1	<.001	1.0, 1.1
% foreign born	1.1	<.001	1.0, 1.1	*not included*			1.0	0.02	1.0, 1.1
% not hs grad	*not included*			1.0	0.01	1.0, 1.0	*not included*		
% no English spoken	0.8	<.001	0.8, 0.9	0.9	0.04	0.7, 0.8	0.8	<.001	0.8, 0.9
% below poverty	1.0	0.00	1.0, 1.0	1.0	0.03	1.0, 1.0	1.0	0.0	1.0, 1.0
County Level									
% ever received sigmoidoscopy/colonscopy	0.8	<.001	0.8, 0.8	0.8	<.001	0.8, 0.9	0.8	<.001	0.8, 0.8
% received fobt last 2 years	*not included*			1.0	0.03	0.9, 1.0	1.0	0.01	0.9, 0.9

Table 3c. Heirarchical, area-based risk models for non-Hispanic whites, by method of handling unknowns and Model Type

Conclusions

- Currently no standard method on how to handle cases with unknown stage at diagnosis
- Most remove from analysis
  - Results from this method deviated the most from the other two
  - Reduces power due to reduce n
  - Creates geographic selection bias; can overestimate effects
- Recoding to late/allocation
  - Potential for misclassification bias; more likely to move results to null

Moving forward…

- Re-abstraction studies or links with clinical datasets
  - To determine a more precise allocation method
- Multiple imputation
  - Compare to other methods
- Results consistent among multiple methods can be interpreted with more confidence

Acknowledgments

PhD dissertation in Epidemiology

Dr. David Lee, Public Health Sciences, Miller School of Medicine, University of Miami

Dr. Kevin Henry, Department of Epidemiology, Rutgers University

Questions?

Recinda Sherman

rsherman@med.miami.edu