

Assessment of Duplicate Cancer Cases in Utah and Idaho: Improving Interstate Cancer Surveillance

Chris Johnson, MPH

Cancer Data Registry of Idaho

NAACCR 2013 CONFERENCE

June 13, 2013

Cancer Data
Registry of Idaho



The Team

- Christopher Johnson, MPH, Cancer Data Registry of Idaho
- Kimberly A. Herget, MStat, Utah Cancer Registry
- Rosemary Dibble, CTR, Utah Cancer Registry
- Stacy Carson, RHIT, CTR, Cancer Data Registry of Idaho
- Antoinette M. Stroup, PhD, Utah Cancer Registry

The Project

- Cancer surveillance across the United States (US) occurs within state and territorial administrative boundaries.
- Central registry staff make decisions about residential status and reportability based on information submitted by health care providers and hospital tumor registrars.

The Project

- Because these data are de-identified and pooled for national statistics, it is not possible to consolidate information on the same case from multiple states and address issues related to case duplication.
- Furthermore, it is unknown what impact interstate consolidation and de-duplication might have on measures of the US cancer burden, such as incidence rates, or data quality indicators, such as percent of cases reported solely via death certificates.

The Project

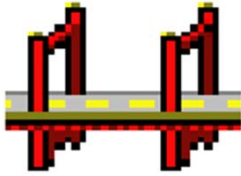


Cancer Data
Registry of Idaho



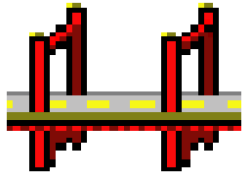
IRB Determinations

- Submitted to both University of Utah and Idaho Division of Public Health Institutional Review Boards.
- Determination by both IRBs was that project did not meet the definitions of Human Subjects Research according to Federal regulations.



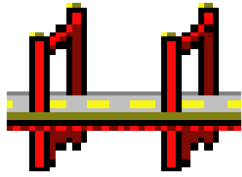
Person Linkage

- Each state created a dataset of 1980-2010 state resident cases.
- From these, each state created a person-specific dataset for the linkage using the most recently reported case for each person.
- Utah: 174,571 people
- Idaho: 136,125 people



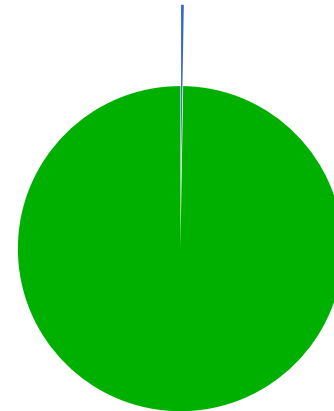
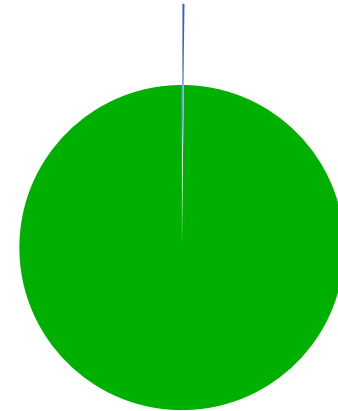
Person Linkage

- **Link Plus Software**
 - (V2 on one machine, V3 on another)
- **Blocking Variables:**
 - Soundex functions of Last Name, First Name, Middle Name; SSN, Birth Date.
- **Matching Variables:**
 - Last Name, First Name, Middle Name, SSN, Birth Date, Sex.
- **Manual Review.....**
 - With High Cutoff Score = 20.0, **N=344 matches**
 - After Manual Review, **N=352 matches**
 - Lowest Score deemed to be a match = 13.9



Person Linkage

- N=352 matches
- Utah:
 $352 / 174,571 = 0.20\%$
- Idaho:
 $352 / 136,125 = 0.26\%$



Case Deduplication

- For the 352 matching people, collect all of the cases from each registry.
 - Multiple records per person – one for each tumor.
- Utah: 381 cases 1980-2010
2 cases earlier than 1980
2 cases later than 2010
- Idaho: 386 cases 1980-2010
- Combined: **771 cases**

Adjudicate Potential Duplicates

- How to decide which state keeps the cases?



Adjudicate Potential Duplicates

1. Using IARC/IACR MP/H Rules (IARCcrgTools).
2. Using alpha version of SAS Program to implement SEER MP/H Rules.
3. CTR staff review case history for each person and use SEER MP/H rules to determine duplicate cases.

Adjudicate Potential Duplicates Using IARC MP/H Rules

- Three runs of IARCcrgTools Multiple Primary Program
 - Idaho cases alone
 - Utah cases alone
 - Combined cases

Adjudicate Potential Duplicates Using IARC MP/H Rules

1995-2010 Case Counts for Linked Persons
Under IARC MPH Rules

Type of Reporting Source	One State Alone		Combined Data		Characteristics of "Double-Counted"	
	Idaho	Utah	Idaho	Utah	Idaho	Utah
Hospital inpatient	156	189	109 70%	157 83%	47 30%	32 17%
Laboratory only	10	24	6 60%	18 75%	4 40%	6 25%
Physician office	31	3	18 58%	1 33%	13 42%	2 67%
Death certificate only	35	5	7 20%	1 20%	28 80%	4 80%
Other	1	0	1 100%	0 -	0 0%	0 -
Total	233	221	141 61%	177 80%	92 39%	44 20%

Note that IARC MPH Rules do not consider in situ cases besides bladder.

Adjudicate Potential Duplicates Using IARC MP/H Rules

1995-2010 Case Counts for Linked Persons
Under IARC MPH Rules

Primary Site (Top 10)	One State Alone					
	Idaho			Utah		
	Beh 0	Beh 2	Beh 3	Beh 0	Beh 2	Beh 3
Prostate	0	0	43	0	0	41
Breast	0	0	31	0	0	26
Lung and Bronchus	0	0	19	0	0	18
Melanomas - Skin	0	0	19	0	0	10
Colon and Rectum	0	0	14	0	0	15
Leukemia	0	0	14	0	0	14
Non-Hodgkin Lymphoma	0	0	10	0	0	14
Brain and Other Nervous System	1	0	9	4	0	6
Urinary Bladder	0	3	6	0	7	6
Pancreas	0	0	8	0	0	7

Note that IARC MPH Rules do not consider in situ cases besides bladder.

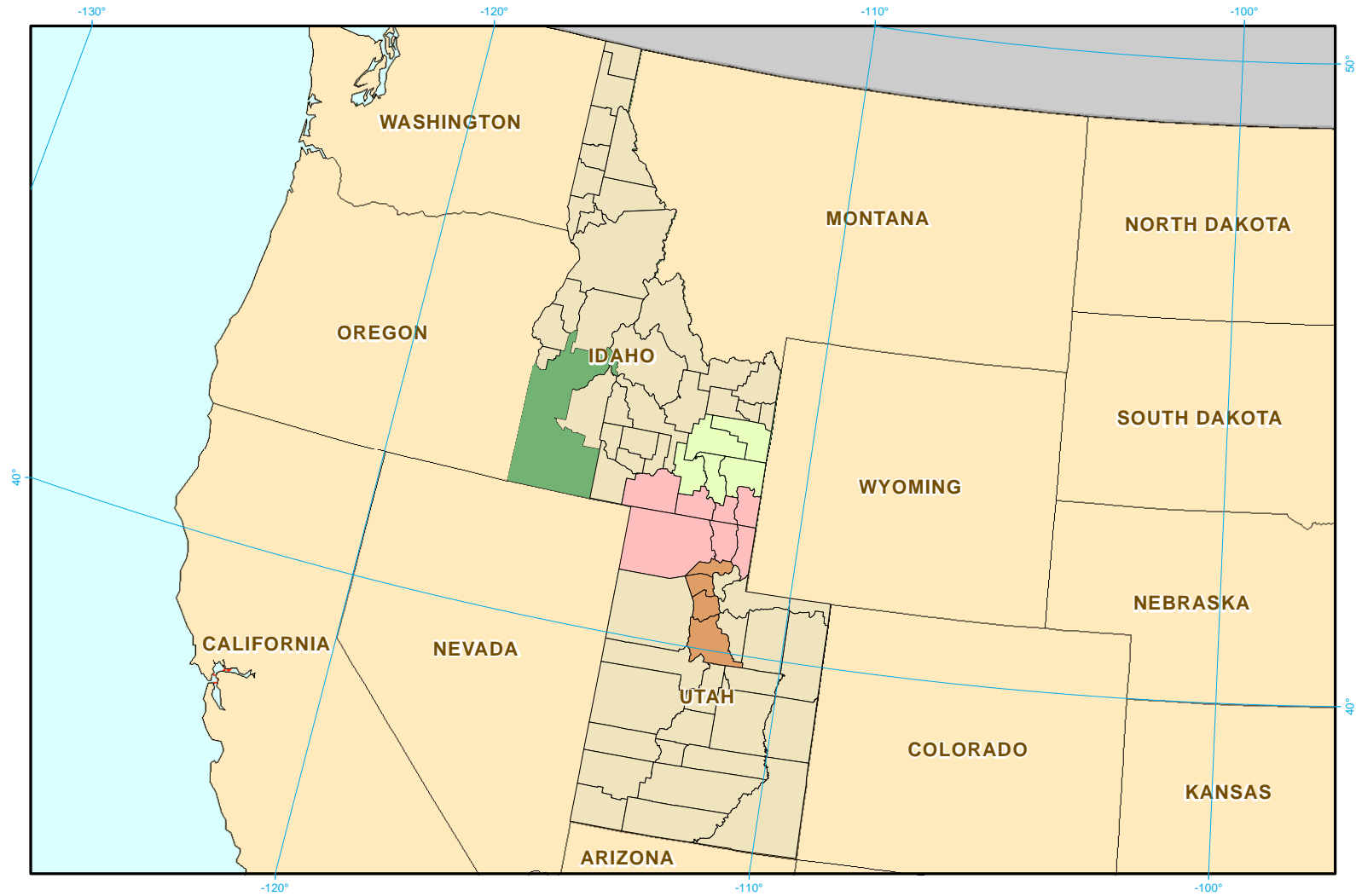
Adjudicate Potential Duplicates Using IARC MP/H Rules

1995-2010 Case Counts for Linked Persons
Under IARC MPH Rules

Primary Site (Top 10)	Characteristics of "Double-Counted"					
	Idaho		Utah			
	Behavior 3		Behavior 2		Behavior 3	
Prostate	22	51%	0		11	27%
Breast	17	55%	0		6	23%
Lung and Bronchus	5	26%	0		4	22%
Melanomas - Skin	7	37%	0		1	10%
Colon and Rectum	4	29%	0		4	27%
Leukemia	6	43%	0		3	21%
Non-Hodgkin Lymphoma	4	40%	0		1	7%
Brain and Other Nervous System	4	44%	0		0	0%
Urinary Bladder	3	50%	2	29%	1	17%
Pancreas	3	38%	0		0	0%

Note that IARC MPH Rules do not consider in situ cases besides bladder.

Does Geography Matter?



Cancer Data
Registry of Idaho



Adjudicate Potential Duplicates Using IARC MP/H Rules

1995-2010 Case Counts for Linked Persons
Under IARC MPH Rules

Geographic Area	State Alone	"Double-Counted"	
Idaho: Border	35	11	31%
Idaho: Near Border	74	32	43%
Idaho: Boise Metro	54	19	35%
Idaho: Other	70	30	43%
Utah: Border	25	3	12%
Utah: Wasatch Front	142	26	18%
Utah: Other	54	15	28%

Note that IARC MPH Rules do not consider in situ cases besides bladder.

Adjudicate Potential Duplicates Using SEER MP/H Rules (SAS program)

- Actually, 19 SAS programs that create lookup tables, define functions, and run macros to process the cases.
- Thanks!!!
 - Gary Levin, Florida Cancer Data System and Florida Department of Health
 - Genevieve Boucher, Quebec Public Health Information Center
 - Rabia Louchini, Quebec Public Health Institute

Adjudicate Potential Duplicates Using SEER MP/H Rules (SAS program)

1995-2010 Case Counts for Linked Persons
Under SEER MPH Rules

Type of Reporting Source	One State Alone		Combined Data		Characteristics of "Double-Counted"	
	Idaho	Utah	Idaho	Utah	Idaho	Utah
Hospital inpatient	166	202	130 78%	181 90%	36 22%	21 10%
Laboratory only	12	28	8 67%	24 86%	4 33%	4 14%
Physician office	36	3	28 78%	1 33%	8 22%	2 67%
Death certificate only	35	5	17 49%	2 40%	18 51%	3 60%
Other	1	0	1 100%	0 -	0 0%	0 -
Total	250	238	184 74%	208 87%	66 26%	30 13%

Adjudicate Potential Duplicates Using SEER MP/H Rules (SAS program)

1995-2010 Case Counts for Linked Persons
Under SEER MPH Rules

Primary Site (Top 10)	One State Alone					
	Idaho			Utah		
	Beh 0	Beh 2	Beh 3	Beh 0	Beh 2	Beh 3
Prostate	0	0	43	0	0	41
Breast	0	3	31	0	5	27
Melanomas - Skin	0	6	21	0	5	12
Lung and Bronchus	0	0	19	0	0	19
Colon and Rectum	0	0	15	0	0	15
Leukemia	0	0	15	0	0	14
Non-Hodgkin Lymphoma	0	0	10	0	0	14
Brain and Other Nervous System	1	0	9	4	0	6
Urinary Bladder	0	3	6	0	7	6
Pancreas	0	0	8	0	0	7

Adjudicate Potential Duplicates Using SEER MP/H Rules (SAS program)

1995-2010 Case Counts for Linked Persons
Under SEER MPH Rules

Primary Site (Top 10)	Characteristics of "Double-Counted"			
	Idaho		Utah	
	Beh 2	Beh 3	Beh 2	Beh 3
Prostate	0	22 51%	0	11 27%
Breast	1 33%	7 23%	0 0	2 7%
Melanomas - Skin	0 0%	1 5%	0 0	0 0%
Lung and Bronchus	0	4 21%	0	5 26%
Colon and Rectum	0	5 33%	0	1 7%
Leukemia	0	4 27%	0	1 7%
Non-Hodgkin Lymphoma	0	4 40%	0	0 0%
Brain and Other Nervous System	0	4 44%	0	0 0%
Urinary Bladder	0 0%	2 33%	2 29%	1 17%
Pancreas	0	3 38%	0	0 0%

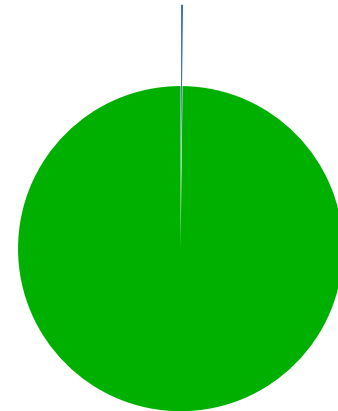
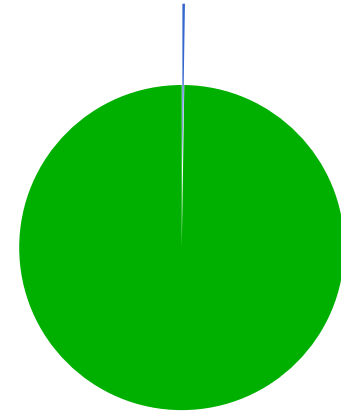
Adjudicate Potential Duplicates Using SEER MP/H Rules (SAS program)

1995-2010 Case Counts for Linked Persons
Under SEER MPH Rules

Geographic Area	State Alone	"Double-Counted"	
Idaho: Border	36	8	22%
Idaho: Near Border	79	28	35%
Idaho: Boise Metro	59	9	15%
Idaho: Other	76	21	28%
Utah: Border	27	1	4%
Utah: Wasatch Front	155	19	12%
Utah: Other	56	10	18%

Duplicate Case Adjudication

- Under SEER Rules:
- Idaho, 1995-2010:
 - $66 / 98,981 = 0.07\%$
 - $250 / 98,981 = 0.25\%$
- Utah:
 - $30 / 128,707 = 0.02\%$
 - $238 / 128,707 = 0.18\%$



Limitations and Caveats

- IARCcrgTools does not consider in situ cases besides bladder.
- In the combined incidence dataset, Utah data had 6 digits for date of diagnosis, so there is some fuzz built in regarding which state, ID or UT, should claim the case.
- The SAS program for implementing SEER MPH rules *REQUIRES* full dates of diagnosis, so these were imputed with day=15 or mmdd=0701. This caused some discrepancies in comparison to the IARC rules, so I hand-edited 15 cases to make them consistent.

The Project

- To our knowledge, this is the first attempt to assess cancer case duplication across state boundaries.

What is left to do?

- One state's DCO is the other state's lost to follow-up
- Accuracy of primary site for DCO cases is not great (often, a metastatic site is listed in place of the primary site).
- Some of the remaining DCO cases are likely duplicates, but the site categories don't match.
- This will only be reconciled by CTRs using full text, potentially death certificates, and may require follow-back to hospitals.

Recommendations for NAACCR

- Although not a significant issue for Utah/Idaho, interstate linkages may yield very different results for populations that are more transient or high "commuting" areas such as NY/NJ, NJ/PA, MD/DC etc.
- There may be more bang for the buck in focusing on DCO cases.
 - In Idaho, 18-35 DCOs out of 2,200 for 1995-2010 could "go away" (up to 1.6%).
 - Could have a significant impact in registries with high DCO rates.

