NAACCR 2013 Conference
THINKING BIG: The Future of Cancer Surveillance
Austin, Texas
June 8 – 14, 2013
The Institute of Medicine and the National Cancer Institute have encouraged researchers to utilize health data set linkages to broaden research perspectives that might inform efforts to improve the quality of cancer care and reduce disparities.
Linkages of Central Cancer Registry Data and Health-Related Data Sets Enhance Study of Breast Cancer Disparities

Therese A. Dolecek
Firas Dabbous
Jenna A. Khan
Jennifer Orsi
Seijeoung Kim
Garth H. Rauscher
The University of Illinois at Chicago
National Cancer Institute P50 CA106740

Focus: Breast Cancer Disparities

Richard B. Warnecke, Ph.D.
Principal Investigator
Breast Cancer Mortality Rates
Chicago (1980-2005)

Hirschman et al. (2007) Cancer Causes and Control
Overview

4 Examples of Linkage Research
Linking Central Cancer Registry Data to:

1. 1990 and 2000 Census Data
2. Illinois Hospital Discharge Data
3. Greater Circle Distance Measures
4. Radiology Data
CPHHD Project: Neighborhood and Individual Effects on Stage at Diagnosis

CPHHD Investigators:
Richard T. Campbell, Ph.D.
Richard E. Barrett, Ph.D.
Therese A. Dolecek, Ph.D.
Richard B. Warnecke, Ph.D.
Linkage to 1990 and 2000 Census Data

CPHHD Project: Neighborhood and Individual Effects on Stage at Diagnosis

1994-2000 Female Breast Cancer Incidence Data in Cook County, Illinois

Supplement Study: 1998-2002
Boston, Chicago (updated), Detroit,
Los Angeles, Philadelphia, Washington DC
Percent Stage at Diagnosis of Breast Cancer by Race/Ethnicity, Cook County, Illinois, 1994-2000

Research Aims

• Examine effects of race, ethnicity, SES and age on stage at diagnosis
• Focus on disparities both in terms of race/ethnicity and SES
• Explore race-SES interactions
• Communicate results to the policy community and the community at large.
ISCR Data, 1994-2000
Cook County, Illinois

- Female breast cancer cases
- Ages (30-89 years)
- Race/ethnicity (white non-Hispanic, black non-Hispanic, Hispanic)
- Geocoded residential address (census tract, block group, block)
- SEER general summary stage
  - in situ, localized, regional, distant, unstaged
Area-Based Socioeconomic Measure

- Poverty estimates (i.e., the proportion of persons below the poverty line) for residential census tract.
- Census estimates were interpolated to the midpoint of the observation interval (1997) based on the 1990 and 2000 Census.
- The census tract specific female poverty rate for the age and race/ethnic group of the breast cancer case under study.
Final Analytic Data

• Census tract data ($N = 1,137$ tracts) from Cook County, IL, in both 1990 and 2000

• Cancer registry breast cancer data for 21,516 female breast cancer cases residing in these census tracts from 1994-2000
Dependent variable and Covariates

**Dependent variable**
- Ordinal outcome - Stage at diagnosis
  - 1: in situ
  - 2: localized
  - 3: regional
  - 4: distant

**Covariates**
- Age in years (quadratic and cubic effect) age: 30-89
- Poverty (quadratic effect) poverty: 0-85
- Race (white nh, black nh and Hispanic. Coded with two dummy variables.)
- Race/ethnicity by age (age quadratic) interaction
- Race/ethnicity by poverty (poverty quadratic) interaction
Age Effects

Fig. 2. (a) Fitted and observed stage at breast cancer diagnosis by age for women aged 30–89. *Non-Hispanic black (red lines) and non-Hispanic white (black lines), poverty held at median for each group. (b) Fitted and observed stage at breast cancer diagnosis by age for women aged 30–89. Hispanic (red lines) and non-Hispanic whites (black lines), poverty held at median for each group.
Fig. 3. (a) Fitted and observed stage at breast cancer diagnosis by poverty level for women aged 30-89. *Non-Hispanic black (red lines) and non-Hispanic white (black lines), poverty held at median for each group. (b) Fitted and observed stage at breast cancer diagnosis by poverty level for women aged 30-89. *Hispanic (red lines) and non-Hispanic whites (black lines), poverty held at median for each group.
Summary for Six Geographic Areas

- Blacks at greater risk of late stage diagnosis.
- Late stage diagnosis experienced by black women is conditional on age, with the disparity being greatest at younger ages.
- In areas with sufficiently large Hispanic populations, findings similar to blacks though less extreme result for Hispanic women.
- Poverty was found to have a strong effect on the probability of being diagnosed at a later stage, regardless of race/ethnicity.
- So, equal access to early detection and treatment should produce equal outcomes in stage at diagnosis and mortality.
Pilot Study Evaluation of Breast Cancer Disparities Using Large Data Set Linkage

Investigators:
Therese A. Dolecek, Ph.D.
Richard T. Campbell, Ph.D.
Garth H. Rauscher, Ph.D.

Ticket for the Cure
Funding from Lottery Ticket Sales in Illinois
ISCR
Illinois State Cancer Registry
Breast Cancer Incidence
2002-2005

DISTANCE
Mammography Center

AREA
2000 Decennial Census Data

IHDD
Illinois Hospital Discharge Data
Inpatient and Outpatient Records
Breast Cancer Diagnoses
2002-2005
Ticket for the Cure Study:
Pilot Study Evaluation of Breast Cancer Disparities
Using Large Data Set Linkage

<table>
<thead>
<tr>
<th>Linkage Variables</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Illinois State Cancer Registry</strong></td>
<td></td>
<td><strong>Illinois Hospital Discharge Data</strong></td>
</tr>
<tr>
<td><strong>Primary Site</strong></td>
<td></td>
<td><strong>Diagnosis/Related Procedures</strong></td>
</tr>
<tr>
<td><strong>Sex: Females only</strong></td>
<td></td>
<td><strong>Patient Sex</strong></td>
</tr>
<tr>
<td><strong>Reporting Facility Name/Number</strong></td>
<td></td>
<td><strong>Reporting Facility Name/Number</strong></td>
</tr>
<tr>
<td><strong>Diagnosis Date: Month/Day/Year</strong></td>
<td></td>
<td><strong>Date of Admission</strong></td>
</tr>
<tr>
<td><strong>Date of Birth: Month/Day/Year</strong></td>
<td></td>
<td><strong>Date of Discharge</strong></td>
</tr>
<tr>
<td><strong>Case Residential Zipcode</strong></td>
<td></td>
<td><strong>Patient Residential Zipcode</strong></td>
</tr>
<tr>
<td><strong>Case Residential County</strong></td>
<td></td>
<td><strong>Patient Origin: County code</strong></td>
</tr>
</tbody>
</table>
Ticket for the Cure Study:
Pilot Study Evaluation of Breast Cancer Disparities Using Large Data Set Linkage

<table>
<thead>
<tr>
<th>Additional Variables</th>
<th>Illinois State Cancer Registry</th>
<th>Illinois Hospital Discharge Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Race (Detailed)</td>
<td></td>
<td>Secondary Diagnosis Codes (7)</td>
</tr>
<tr>
<td>Computed Hispanic Ethnicity</td>
<td></td>
<td>Principal Procedure Code</td>
</tr>
<tr>
<td>Computed Age at Diagnosis</td>
<td></td>
<td>Secondary Procedure Codes (5)</td>
</tr>
<tr>
<td>Birthplace</td>
<td></td>
<td>Charges (10 categories)</td>
</tr>
<tr>
<td>Case Residence Geocodes</td>
<td></td>
<td>Payer ID/Type Code (3 Possible)</td>
</tr>
<tr>
<td>Latitude/Longitude Coordinates</td>
<td></td>
<td>In Patient Length of Stay (Days)</td>
</tr>
<tr>
<td>Census Tract, Block Group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEER General Summary Stage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morphology</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Linkage

• 38,247 ISCR records - primary breast cancer site
• 161,74 IHDD breast cancer diagnosis
  (subset of 2,961,063 inpatient & 2,961,685 outpatient surgery discharge records for females

Probabilistic Linkage performed Using Automatch software, Matchware Technologies, Inc.

• 29,381 ISCR records linked to a one-to-many match with 44,696 IHDD records
<table>
<thead>
<tr>
<th>Procedures</th>
<th>History Only</th>
<th>Primary Dx</th>
<th>Secondary Dx</th>
<th>Primary or Secondary Dx</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>0.9%</td>
<td>31.5%</td>
<td>7.9%</td>
<td>17.0%</td>
</tr>
<tr>
<td>Biopsy only</td>
<td>7.8%</td>
<td>84.2%</td>
<td>61.2%</td>
<td>82.4%</td>
</tr>
<tr>
<td>Biopsy + Lumpectomy</td>
<td>4.5%</td>
<td>83.1%</td>
<td>57.7%</td>
<td>82.7%</td>
</tr>
<tr>
<td>Biopsy + Mastectomy</td>
<td>0 of 3 records</td>
<td>78.9%</td>
<td>35.9%</td>
<td>76.4%</td>
</tr>
<tr>
<td>Biopsy + Lumpectomy + Mastectomy</td>
<td>0 of 1 record</td>
<td>81.6%</td>
<td>0 of 0 records</td>
<td>81.6%</td>
</tr>
<tr>
<td>Lumpectomy only</td>
<td>5.3%</td>
<td>78.1%</td>
<td>49.4%</td>
<td>77.3%</td>
</tr>
<tr>
<td>Lumpectomy + Mastectomy</td>
<td>0 of 6 records</td>
<td>74.3%</td>
<td>64.0%</td>
<td>73.5%</td>
</tr>
<tr>
<td>Mastectomy only</td>
<td>4.0%</td>
<td>65.0%</td>
<td>58.8%</td>
<td>64.9%</td>
</tr>
<tr>
<td>Any Above Procedure</td>
<td>5.6%</td>
<td>76.4%</td>
<td>55.5%</td>
<td><strong>75.7%</strong></td>
</tr>
</tbody>
</table>
Multiple logistic regression analysis of risk factors and late stage diagnosis of female breast cancer cases (N=5,486) in rural counties of Illinois, 2002-2005

<table>
<thead>
<tr>
<th>Factor</th>
<th>Odds ratio (95% CI)</th>
<th>Factor</th>
<th>Odds ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (yrs)</td>
<td></td>
<td>P</td>
<td></td>
</tr>
<tr>
<td>50-64</td>
<td>Referent</td>
<td>&lt;0.001</td>
<td>Insured</td>
</tr>
<tr>
<td>&lt; 40</td>
<td>2.14 (1.59-2.87)</td>
<td></td>
<td>Medicare</td>
</tr>
<tr>
<td>40-49</td>
<td>1.09 (0.91-1.31)</td>
<td></td>
<td>Medicaid</td>
</tr>
<tr>
<td>65-74</td>
<td>0.78 (0.59-1.04)</td>
<td></td>
<td>Uninsured</td>
</tr>
<tr>
<td>75+</td>
<td>0.98 (0.74-1.30)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td>P</td>
<td>Distance ‡</td>
</tr>
<tr>
<td>White</td>
<td>Referent</td>
<td>&lt;0.01</td>
<td>&lt; 5 miles</td>
</tr>
<tr>
<td>Black</td>
<td>1.92 (1.28-2.89)</td>
<td></td>
<td>5-9 miles</td>
</tr>
<tr>
<td>Other</td>
<td>0.87 (0.362-2.09)</td>
<td></td>
<td>10-14 miles</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>15+ miles</td>
</tr>
</tbody>
</table>

‡ distance from case residence to nearest mammography centers
1 R01 HS018366-01A1
Comparative effectiveness of breast imaging modalities: a natural experiment

Garth H. Rauscher, Ph.D., Principal Investigator
Therese A. Dolecek, Ph.D., Co-Investigator

Purpose is to examine the comparative effectiveness of screening and diagnostic imaging and biopsy procedures.
Specific Aims Overview

- Examine the comparative effectiveness
  • Screening with digital mammography vs. film screen mammography, overall and within clinically relevant subgroups
  • Diagnostic imaging (digital mammography, film screen mammography, breast ultrasound and breast MRI)
  • Diagnostic breast biopsy procedures (stereotactic, ultrasound and MRI-guided needle biopsies; stereotactic vacuum-assisted biopsy and surgical biopsy)
- Describe learning curves associated with the introduction of computer-aided diagnosis and digital mammography
Sources and Methods

• Advocate Health Care
  – PendRad radiology database
    • 355,000 patients
      (~65,000 AA, 23,000 Hispanic)
• Illinois State cancer Registry (ISCR)
  – Incident cancer cases
• Probabilistic linkage (Automatch) on names, DOB, SSN, and weaker identifiers
• Minimize sharing of identifiers
Process for data linkage and creation of analysis datasets

- Vital Status File
- Geocode File
- Raw radiology analysis file
- Raw radiology identifier file
- Breast Cancer Incidence File
- Geocode File
- Raw match file (cancer incidence subsetted to AHC patients)
- Cleaning radiology analysis file (radiology analysis variables)

AHC

ISCR

Preliminary analysis dataset

- Screening examinations
- Diagnostic imaging series
- Biopsy series

BCSC Upload
Figure 3. Relationship between data and study aims

**Aim #1 Screening imaging**

<table>
<thead>
<tr>
<th>Cancer Status</th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abnormal</td>
<td>TP</td>
<td>FP</td>
</tr>
<tr>
<td>Normal</td>
<td>FN</td>
<td>TN</td>
</tr>
</tbody>
</table>

**Aim #2 Diagnostic imaging**

<table>
<thead>
<tr>
<th>Cancer Status</th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biopsy</td>
<td>TP</td>
<td>FP</td>
</tr>
<tr>
<td>No Biopsy</td>
<td>FN</td>
<td>TN</td>
</tr>
</tbody>
</table>

**Aim #3 Diagnostic biopsy**

<table>
<thead>
<tr>
<th>Cancer Status</th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancer</td>
<td>TP</td>
<td>FP</td>
</tr>
<tr>
<td>Normal</td>
<td>FN</td>
<td>TN</td>
</tr>
</tbody>
</table>

Note: BIRADS 3 imaging results will be analyzed separately.
Results of Linkage, April 2013

• 254,155 ISCR records – female primary breast cancer site diagnosed 1/1/1986 – 12/31/2010
• 413,086 AHC radiology records for screening and dx breast imaging exams for females 1/1/2001 – 12/31/2011)

Probabilistic Linkage performed Using Automatch software, Matchware Technologies, Inc.

• 22,741 ISCR records matched in a one-to-many relationship to 20,945 unique AHC records
Coding from the Linkage Algorithm

```plaintext
PROGRAM GEOMATCH
DICTA iscr
DICTB adv
;
BLOCK1 CHAR L_sdx L_SDX
BLOCK1 CHAR F_sdx F_SDX
;
MATCH1 UNCERT LName LAST_NAME 0.9 0.01 700
MATCH1 UNCERT FName FIRST_NAME 0.9 0.01 700
MATCH1 CHAR MidInit MIDDLE_INIT 0.9 0.01
MATCH1 CNT DIFF SSN1 SSN1 0.9 0.01 2
MATCH1 CNT_DIFF DOB2 DOB 0.9 0.01 1
MATCH1 CNT_DIFF ZIP5 ZIP5 0.9 0.01 1
;
CUTOFF1 35 28 35
;
;
BLOCK2 NUMERIC DOB2 DOB
;...
```
Match Rate

The match rate for the 2001-2009 AHC exams was estimated at 98%. Matches for 2010 AHC exam records were also found, but at a much lower match rate, so these data should be used cautiously.
Summary

Population-based central cancer registries play an important role in the study of health disparities and linkages to other health datasets expand the utility of this valuable resource in cancer control and population science.
Acknowledgement
Central Cancer Registries

**Major Source:**
Illinois State Cancer Registry

**Supplement Study Data:**
California Cancer Registry
District of Columbia Cancer Registry
Massachusetts Cancer Registry
Metropolitan Detroit Cancer Surveillance System
Pennsylvania Cancer Registry
Citation