The Integrated Cancer Information and Surveillance System (ICISS)

Team – Systems – Data – Methods

UCRF Initiative Review
April 29, 2014

Anne-Marie Meyer: ameyer@unc.edu
Research Assistant Professor, Department of Epidemiology
Facility Director, ICISS
Web - http://iciss.unc.edu
What is ICISS?

**Scope**

1. To develop a “big data” resource of linked, population-based data across North Carolina.
   - Foundation of NC cancer registry linked to healthcare administrative claims data (public & private)
2. To extend research methods and systems to enable more efficient, systematic, innovative, and rigorous population-science big data research.

**Importance and Relevance**

ICISS provides a unique data resource which helps researchers understand the health of the state population and facilitates the goal of the UCRF to save lives and reduce suffering from cancer.

- The ICISS data set North Carolina apart in its in its ability link individuals across multiple insurance sources, registries, and patient cohorts.
- The interdisciplinary team are able to apply diverse and novel methods to answer important population-based questions that span the cancer continuum.
- Innovative systems have been developed to support the entire research lifecycle, have enabled better science, have been adopted by other research groups and UNC computing.

ICISS is directly able to assess the impact of state and federal health initiatives (e.g., the ACA).
ICISS Accomplishments

• Enabled **26 peer-reviewed manuscripts**, several in high impact medical and specialty journals.
  – 4 in *Cancer*; 4 in *Medical Care*; 2 in *JAMA*

• Facilitated **$10 million in grant funding:**
  – 16 grants total (direct involvement)
    – $7.5 million in funding
    – $7.1 million **external**
    – $380K **internal**
  – 8 grants awarded using ICISS Clinical Coding System ($2.7 million)

• Created an innovative software tool which has been highlighted at national conferences and **licensed through UNC Office of Technology Development**.

• Collaborated across **4 UNC schools, 16 departments, 6 other UCRF initiatives**.

• Provided a strong foundation for the research careers of **19 junior faculty** and **16 post-doctoral trainees and students**.
## Research Productivity

<table>
<thead>
<tr>
<th>Top Three</th>
<th>Papers</th>
<th>Grants</th>
</tr>
</thead>
</table>
  – Sponsor: AHRQ  
  – Total: $1.1 million  
  – PI: Alan Brookhart |
  – Sponsor: American Cancer Society  
  – Total: $727,000  
  – PI: Stephanie Wheeler |
  – Sponsor: Robert Wood Johnson Foundation  
  – Total: $199,876  
  – PI: Anna Schenck / Anne-Marie Meyer |

| Total Attributable to UCRF | 26 published manuscripts, including high impact and specialty journals  
  • 4 in *Cancer*  
  • 4 in *Medical Care*  
  • 2 (+1 letter) in *JAMA* | 16 grants total  
  • $7.5 million in funding  
  – $7.1 million external  
  – $380K internal |
|                          | Current manuscript pipeline:  
  • 6 in press  
  • 8 submitted or in revision  
  • 14 in progress | 8 grants won using ICISS Clinical Coding System ($2.7 million) |

| Current grant pipeline: | 6 awaiting award decision (including 2 R01s)  
  • 3 in preparation for spring submission (including 1 U grant) |
Linked Data for Population Research

**What**: Linkage of North Carolina Central Cancer Registry (NC CCR) cases to claims:

- 5.5 million individuals or 55% of the state’s total population.
- Beneficiaries are reconciled across payers and linked to ~400,000 cancer cases.
- Several high profile publications highlight this effort, including ICISS conceptual model, cancer data ontology, and linking methodologies.

**Why**: To understand burden of disease from a population perspective, one needs “population-based” data.

- More heterogeneous and representative than clinical trials and single-payer claims data (i.e., the “missing patients”).
- Requires deeper contextual knowledge (e.g., time and place).
# How to Understand the Health of a Population

## Fields of Research
- Epidemiology
- Comparative Effectiveness (CER)
- Health Disparities
- Quality Improvement (QI)
- Patient Centered Outcomes Research (PCOR)
- Public Health Systems and Services Research (PHSSR)
- Implementation Science
- “Methods”
- Health Information Technology (HIT)
- Health Economics
- Organizational Studies
- ...

## Substantive Examples
- Patient heterogeneity
- Health care utilization
- Survivorship / Care trajectories
- Guideline concordant care
- Provider and system influences
- Networks (patients & providers)
- Geospatial risk, access, outcomes
- Changes in policy!

These areas of research are of national interest and growing in importance with the implementation of the ACA and the development of PCORnet.

![Figure 2a: ET initiation by Race, chemotherapy](image)
Team Science: Spatial & Social Context


*Number showing in each county are the number of endoscopy facilities.
• IMRT compared with conformal radiation therapy:
  – Less gastrointestinal morbidity
  – Fewer hip fractures
  – More erectile dysfunction

• IMRT compared with proton therapy:
  – Less gastrointestinal morbidity

Some insurance providers are no longer paying for proton therapy.
ICISS Timeline

- Initial linkages, systems and tools development took approximately **2 years** to mature.
  - Single tumor site to single payer.

- Complete data linkage took **18 months** to mature.
  - Entire NC registry to all payers (>2.5 terabytes raw data).

- Parallel systems development was key to increased efficiency and quality.
ICISS Systems Impact

ICISS has developed a portfolio of systems to address the challenges inherent to managing, manipulating, and analyzing Big Data.

- Support the entire research lifecycle.
- Make data more transparent and available.
- Operationalize and manage data governance and security.
- Leveraged by multiple UCRF initiatives and cancer center partners.
- Made a global impact on campus computing through investment of ITS Research Computing into ICISS computing environment (SeDAP).
Aggregated Data System

Map of North Carolina showing mortality rates with specific data for Robeson County:

- **ARF - 2009-2010**: % Below Poverty - 22.8, Rank 4, Range 7.1-23.9
- **No Insurance 2000**: 22.9, Rank 1, Range 10.8-22.9
- **BRFSS - 2008**: Sigmoidoscopy/Colonoscopy / YES - 51.33, Rank 23, Range 51.33-75.73
- **CCR - 2003-2007**: Colon/Rectal Cancer / Total / Rate - 51.8, Rank 36, Range 38-71.6
- **Colon/Rectal Cancer / Total / Mortality Rate**: 20.9, Rank 26, Range 8.4-29.9
- **CHSI - 2009**: Obesity - 31.5, Rank 10, Range 8.3-36.9
- **RWJ - 2010**: Access to Care - 36, Rank 37, Range 0-100
- **Adult Smoking**: 28, Rank 20, Range 0-40.3

* indicates values <10, - indicates missing, # indicates only aggregated for specific locations.
Comprehensive Coding System

### Medication - NDC Download

For generic medication: Atorvastatin

You are about to download content that is licensed to your name. Please do not distribute this data.

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GENERIC_NAME</td>
<td>Name of the generic medication</td>
</tr>
<tr>
<td>DRUG_NAME</td>
<td>Name of the packaged drug</td>
</tr>
<tr>
<td>STATUS</td>
<td>[Active</td>
</tr>
<tr>
<td>ROUTE</td>
<td>Route of administration</td>
</tr>
<tr>
<td>STRENGTH</td>
<td>Strength of individual dose</td>
</tr>
<tr>
<td>STRENGTH_UOM</td>
<td>Unit of measure or strength</td>
</tr>
<tr>
<td>FORM</td>
<td>Form of delivery</td>
</tr>
<tr>
<td>AVAILABILITY</td>
<td>[Prescription (Rx)</td>
</tr>
<tr>
<td>DEA</td>
<td></td>
</tr>
<tr>
<td>NDC</td>
<td></td>
</tr>
<tr>
<td>PACKAGE_SIZE</td>
<td></td>
</tr>
<tr>
<td>APPROVAL_DATE</td>
<td></td>
</tr>
<tr>
<td>TERMINATION_DATE</td>
<td></td>
</tr>
<tr>
<td>MANUFACTURER</td>
<td></td>
</tr>
</tbody>
</table>

Macro Variable `breast_chemo_codes_in` | Remove Punctuation (e.g.: C50.3 = C503)

/* macro variables */

let `breast_chemo_codes (inc biologics) icd9dx` = ('V58.1', 'V58.11', 'V66.2', 'V67.2');
let `breast_chemo_codes (inc biologics) procdc9` = ('99.25');
let `breast_chemo_codes (inc biologics) medname` = ('PDB-MN000000502', 'PDB-MN000003431', 'PDB-MN0000003508', 'PDB-MN0000000995');
let `breast_chemo_codes (inc biologics) prochpcs` = ('94640', '96469', '96410', '96411', '96412', '96413', '96414', '96415');

/* call to macro variables */

if icd9dx in `breast_chemo_codes (inc biologics) icd9dx` then ....
if procdc9 in `breast_chemo_codes (inc biologics) procdc9` then ....
if medname in `breast_chemo_codes (inc biologics) medname` then ....
if prochpcs in `breast_chemo_codes (inc biologics) prochpcs` then ....
Secure Data Analysis Platform (SeDAP)

Built in conjunction with ITS-Research Computing
Team Science Approach

ICISS research partners represent 16 departments across 4 schools and 4 external organizations.

ICISS staff include PhD and Master’s level biostatisticians, demographers, statistical mathematicians, software developers, and database specialists.

Current and potential research partners include RTI International, Georgia State University, University of Colorado, North Carolina State University, Duke, and Wake Forest.
Future Plans

2014

Integrate updated claims data
Develop common data model

2015

Expand linkages to new data domains
Adopt new supporting technologies as needed

2016 – Future

Deploy Knowledge Sharing System
Build data warehouse
Complete SeDAP migration
Deploy Aggregated Data 2.0

Near-Term Highlights:

• Multi-Payer Claims Database Workshop

• ICISS Roadshows Highlighting:
  – SeDAP
  – Data Linkage
  – Methods Potential

• Current grant pipeline:
  – 6 awaiting award decision (including 2 R01s)
  – 3 in preparation for spring submission (including 1 U grant)

Methods

R01 for methods development
Recruit 5 new collaborating PIs, 1 at Associate level

Continuously develop linking algorithms and foster linking methods development
Questions

Anne-Marie Meyer – ameyer@unc.edu
Web – http://iciss.unc.edu
ICISS Staff

ICISS Development & Leadership Team
Andrew Olshan; *Optimizing Outcomes Theme Leader*
Ethan Basch, *Director Cancer Outcomes*
William Carpenter; *Faculty Director*
Anne-Marie Meyer; *Facility Director*
Adrian Meyer; *Director of Systems Development*

Statistics and Data Management
YunKyung Chang, PhD
Tzy-Mey (May) Kuo, MPH, PhD
Ke Meng, PhD
*Bong-Jin Choi, PhD*
*Lei Zhou, MSPH*
*joining this summer*

ICISS System Development
Ciararro Faulk, BS
Jabari Jerkins, MIS
Jonathan Tweedy, MIS

Study Coordination and Project Management
Laura Green, MBA, PMP
## Research Productivity

<table>
<thead>
<tr>
<th>Impact of UCRF Investment to Date</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Recruitments</strong></td>
</tr>
<tr>
<td>Ethan Basch</td>
</tr>
<tr>
<td>Jennifer Lund</td>
</tr>
<tr>
<td>Anne-Marie Meyer</td>
</tr>
<tr>
<td>Matthew Nielsen</td>
</tr>
<tr>
<td>Angie Smith</td>
</tr>
<tr>
<td>Karyn Stitzenberg</td>
</tr>
<tr>
<td>Justin Trogdon</td>
</tr>
<tr>
<td><strong>Key UNC Collaborators</strong></td>
</tr>
<tr>
<td>ICISS has provided a strong foundation for the research careers of 19 junior faculty and 16 post-doctoral students and trainees.</td>
</tr>
<tr>
<td>• 13 of these junior faculty are UCRF recruits</td>
</tr>
<tr>
<td>UNC collaborators from:</td>
</tr>
<tr>
<td>• SOM, SON, SPH, SILS, RENCI, ITS-Research</td>
</tr>
<tr>
<td>• 16 departments</td>
</tr>
<tr>
<td>• 6 UCRF initiatives (CBCS, Health Registry, Health eNC/Chart, UNCSSeq, Bioinformatics, Clinical Trials)</td>
</tr>
<tr>
<td><strong>Key External Collaborators</strong></td>
</tr>
<tr>
<td>ICISS has begun teaming with external organizations, including:</td>
</tr>
<tr>
<td>• RTI International – data and analytic support</td>
</tr>
<tr>
<td>• NC State – data</td>
</tr>
<tr>
<td>• University of Colorado – data and analytic support</td>
</tr>
<tr>
<td>• Georgia State University – analytic support</td>
</tr>
</tbody>
</table>
## Major Milestones

### Data

- **2010**: Pilot linkage complete (single site, single payer)
- **2011**: CCR cases & claims data acquired
- **2012**: Initial linkage complete
- **2013**: Medicare Part D claims acquired
- **2014**: Claims data update

*Develop a linked data resource of large, population-level datasets*

### Systems

- **2010**: Clinical Coding & Tracking Systems launched
- **2011**: SeDAP approved
- **2012**: Clinical Coding System presented at AcademyHealth
- **2013**: SeDAP Beta Go-Live

*Develop critical systems and systems infrastructure*

### Methods

- **2010**: IRB approved
- **2011**: CER data framework paper published
- **2012**: JAMA paper published / CER data survey & ontology published
- **2013**: First manuscript published using initial linkage
- **2014**: Nine manuscripts published in 2013

*Facilitate cutting edge cancer outcomes research*

### Operations

- **2010**: Steering Committee established
- **2011**: Governance Charter approved
- **2012**: Security mgmt program implemented
- **2013**: Governance Charter revised
- **2014**: Security mgmt program updated

*Operationalize the resource*
ICISS Strategic Framework

**Strategic Goals**
- Build and maintain a comprehensive linked data resource
- Deliver systems and tools to support big data research
- Facilitate cutting edge cancer outcomes research

**Objectives**
- Obtain and maintain a comprehensive, novel, and longitudinal data resource
- Develop a common data model
- Develop validated linkages within the data
- Operate a secure data analysis platform (SeDAP)
- Develop IT solutions to enable research operations
- Develop advanced data tools to support the research lifecycle
- Develop methods addressing data, modeling, and systems for secondary data research
- Publish high impact research products
- Secure high value grants

**Operations Management and Support**
Linked Data for Population Research

Geographic/Area Level Data

System and Provider Level Data

Person Level Data

Reference Data

Characterize environment, socioeconomic, and political context
- Census Data
- Robert Wood Johnson County Health Rankings
- CDC County Indicators Data
- Area Resource File
- Environmental Monitoring Data

Practice location, Specialty, Demographic, Treatment Patterns
- NC Health Professionals Data System
- American Medical Association Data
- American Hospital Association
- National Association of County and City Health Officials (NACCHO)

Patient level demographics, risk factors and tumor information
- Cancer Registry Data
- Federal and Private Payer Claims Data
- Other Registries
- Linked Cancer-Related Cohort Data

Normalize and link multiple nomenclatures through crosswalks
- ICD-9, ICD-10
- CPT/HCPCS
- LOINC
- Zipcode
- Census Tract