

# **Strengthening State Cancer Registry Data by Linking to Public and Private Sources**

**Joseph Lipscomb, PhD  
Emory University**

***A Foundation for Evidence-Driven Practice:***

***A Rapid Learning System for Cancer***

**National Cancer Policy Forum**

**Institute of Medicine**

**October 5, 2009**

# *Immediate Question.....*

**In promoting a Learning Health Care System,  
why emphasize:**

- n Registries?**
- n Linking to other data sources?**
- n Cancer?**
- n State-level focus?**

*Note: the IOM's National Cancer Policy Board has urged development of a "national cancer data system"*

\* Hewitt M, Simone JV (eds). *Enhancing Data Systems to Improve the Quality of Cancer Care*. Washington, DC: National Academy Press; 2000.

# *Why Registries?*

- n As a “organized system that uses observational study methods to collect uniform data to evaluate outcomes\* . . . .,” a disease registry is a core resource for a learning health care system.
- n **Potential purposes: describe natural history of disease, determine clinical effectiveness or cost-effectiveness, measure & monitor safety and harm, evaluate quality of care.**
- n **Types: product registries (specific drugs or devices), health services registries (specific clinical procedures, encounters, hospitalizations), and disease registries (specific conditions)**
- n ***An unrealized ideal: a population-based disease registry that can serve both as a health services registry and product registry.***

\* Gliklich RE, Dreyer NA, eds. *Registries for Evaluating Patient Outcomes: A User's Guide*. AHRQ Publication No. 07-EHC001-1. Rockville, MD, April 2007.

# *Why Link to Other Data Sources?*

q Even the best registries may not be adequate for addressing key health system questions, e.g., in comparative effectiveness research (CER) or cost-effectiveness analysis (CEA).

e.g., NCI's Surveillance, Epidemiology and End Results (SEER) program routinely collects patient demographics, primary tumor site, morphology, stage at diagnosis, first course of treatment, and follow-up survival status on each patient. *But not information on treatment beyond first course, disease recurrence, resources consumed, provider characteristics, patient-reported outcomes.....*

q When possible, augment registry data by capitalizing on existing data – avoid collecting anew what already exists somewhere.

q Multiple sources of information on the same event may permit cross-validation to improve data accuracy, e.g., do registry and claims data agree on surgery type and date?

# *Why Cancer?*

## **n State-of-the-art population-representative disease registries**

*SEER*

*National Program of Cancer Registries (NPCR)*

*National Cancer Data Base (NCDB)*

## **q State-of-the-art creation and application of linked data sets to support health services and outcomes research**

*Registry - Administrative/Claims Data*

(e.g., SEER-Medicare; SEER-Medicaid; State Registry-Private Claims)

*Registry – Medical Records – Claims - Provider Descriptors*

(e.g., CDC's Breast and Prostate Data Quality and Patterns of Care Study)

*Registry - Medical Records - Patient Reports*

(e.g., NCI's Prostate Cancer Outcomes Study)

*Registry – Medical Records – Patient/Physician/Caregiver Reports*

(NCI-VA Cancer Care Outcomes Research and Surveillance (CanCORS) Consortium)

# *Why a State-Level Focus?*

As a practical matter, an expeditious route to a **national cancer data system** that is also an effective **health care learning system** may be strong, state-based data systems. Why?

1. **The ever-improving capacity of state registries (benefitting directly and indirectly from NPCR, SEER, and NCDB)**
2. State Comprehensive Control Plans that increasingly call for better state data systems for surveillance and outcomes assessment
3. **The demonstrated capacity to link cancer registry data at the state level with public and private data sources**
4. The reality that – except now for SEER-Medicare – our ability to ***routinely*** link population-based cancer registry data with external administrative or clinical sources to create an integrated multi-state or national system ***starts at the state level*** and arguably ***requires collaboration across states***.
5. **The state may be the right size “laboratory” for learning: large enough to reflect the complexity of multi -data-set linking – small enough still to manage the chaos.**

# *Georgia as One Such Laboratory*

- n **SEER:** 15 GA counties long a part of SEER program
- n **Georgia Central Cancer Registry (GCCR):** focused increasingly, on patterns of care/quality of care studies; “gold” certification from North American Association of Central Cancer Registries
- n **GCCR-Medicare linked data set:** includes all cases diagnosed 1995-2005, with Medicare claims through 2007. Hence, SEER-Medicare link in GA “expands” to include now all 159 counties.
- n **GCCR-Medicaid linked data:** In an American Cancer Society study, Emory investigators collaborated with GA Department of Community Health and contractor Thomson Reuters Healthcare (MedStat) to link GCCR data (1999-2004) to Medicaid claims and enrollment data (1996-2006) to evaluate the Breast and Cervical Cancer Prevention and Treatment Act.

# *Setting the Stage for a New Project....*

## **“Using Cancer Registry Data and Other Sources to Track Measures of Care in Georgia”**

- q 1-year (9/09 – 9/10), \$372,000 grant awarded to Emory jointly by Association of Schools of Public Health and Centers for Disease Control and Prevention, with funding provided by National Cancer Institute and Georgia Cancer Coalition
- q Designed originally as a two-year project, and Year 2 support anticipated now from NCI and possibly other sources
- q Overarching goals
  - Contribute toward development of an integrated, sustainable state-level data system to support cancer research – and learning
  - Support state-level policy objectives (e.g., Data and Metrics objectives in new GA Comprehensive Cancer Control Plan)

## SPECIFIC AIMS

1) For incidence cases of breast and colorectal cancer in Georgia over 1999-2005, link GCCR with administrative and clinical data drawn from the following sources:

- Medicare files

- Medicaid files

- State Health Benefit Plan (State Plan) files covering all GA employees, including public school teachers and their dependents [enrolled in a variety of plans offered by Blue Cross, United HealthCare, Kaiser, others]

- Kaiser Permanente Georgia (KPG) clinical and administrative data

- Medical records and charts (when administrative files inadequate)

- Georgia state hospital discharge data (to capture the uninsured)

- Facility-specific descriptive data (American Hospital Directory)

- Physician-specific descriptive data (the CMS Medicare Physician Identification and Eligibility Record (MPIER) file)

- Other secondary sources (Census data, Area Resource File, geocoding)

*The intent: develop a set of “bilateral” linked, de-identified data sets*

## SPECIFIC AIMS

2) **Subject each bilateral data set to rigorous data quality checks**

3) **Apply each bilateral data set to quality-of-care assessment**

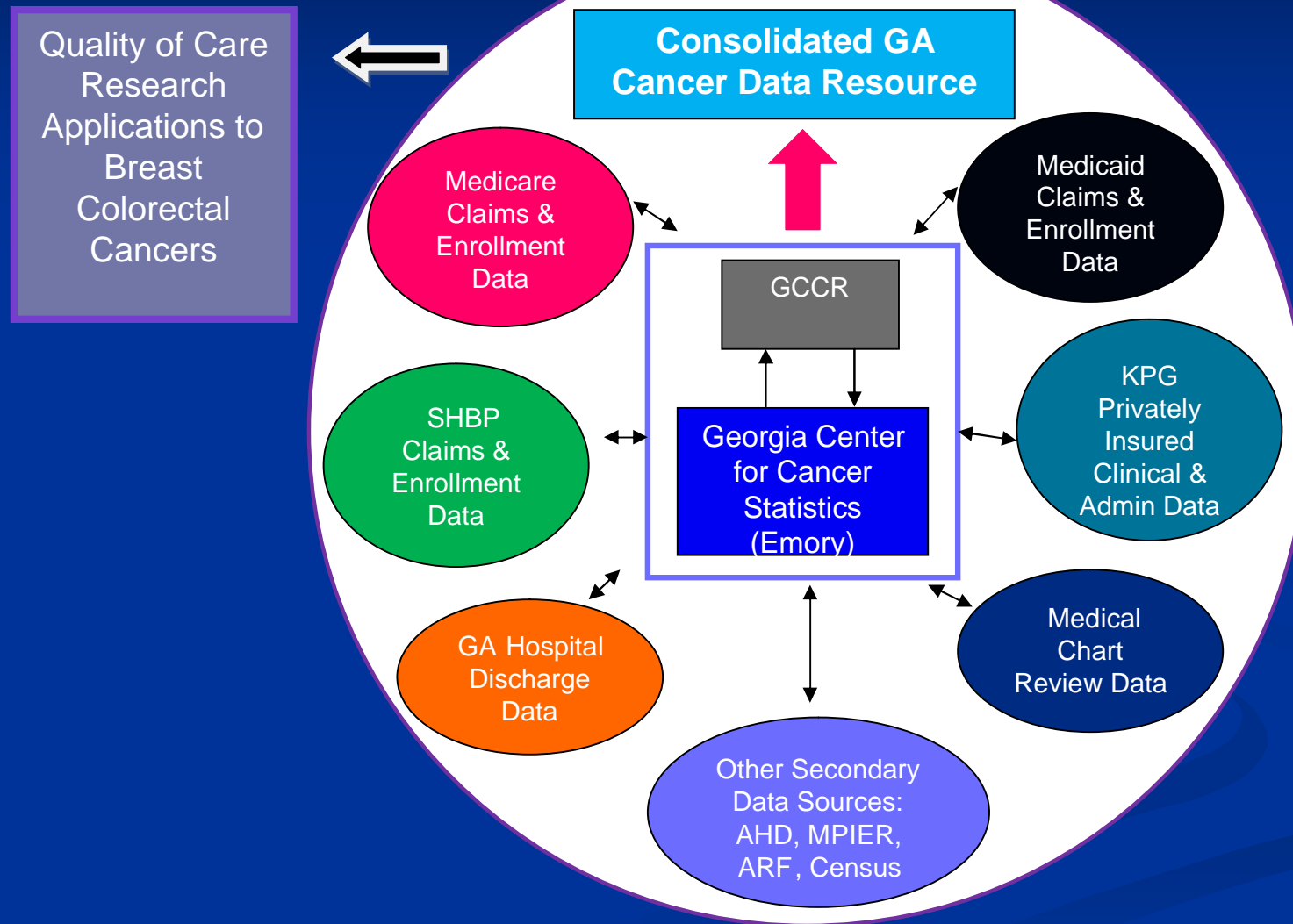
Breast and colorectal cancer quality measures would include those endorsed by the National Quality Forum/ASCO/NCCN; the Institute of Medicine (for the State of Georgia); and others consistent with current guidelines.

4) **Design the alpha version of a “Consolidated Georgia Cancer Data Resource”**

- n Each administrative/clinical data file (which would have its own scrambled ID's) would be linked to GCCR (which would have the crosswalks between each set of scrambled ID's and the patients' SSNs).
- n GCCR could then create linked, de-identified analytical data sets tailored for specific analyses.

*Putting it all together .....*

# Linking GA Cancer Registry Data to Public and Private Sources



**Table 1. Data Elements in the Georgia Comprehensive Cancer Registry (GCCR), Medicare (MCARE), Medicaid (MCAID), State Health Benefit Plan of Georgia (SHBP), Kaiser Permanente of Georgia (KPG), and Georgia Hospital Discharge Data (GA HDD) Files 1999-2005**

Data Elements	GCCR	MCARE	MCAID	SHBP	KPG	GA HDD
Patient ID	X	X	X	X	X	X
DOB	X	X	X	X	X	X
Sex	X	X	X	X	X	X
Race	X	X	X	X	X*	X
Address	X	X*			X	
MCAID ID			X			
SSN	X	X	X	X	X	
Health plan type		X	X	X	X	X
Primary ca site	X				X	
# of cancer sites	X				X	
Date of Dx	X				X	
Cancer Stage	X				X	
Method of Dx	X				X	
Date of death	X	X	X		X*	
Cause of death	X					
Insur. plan detail				X		
MCAID eligibility			X			
MCARE eligibility		X	X		X	
Dates of covrg		X	X		X	
Type of coverage		X	X	X	X	
HMO enroll dates		X	X	X	X	
Diagnostic/ util.					X	
ICD-9 codes		X	X	X	X	X
CPT codes		X	X	X	X	X
Inpt, outpt, & provider services		X	X	X	X	X
Provider ID/Zip		X**	X	X	X	
Pharmacy srvcs.		X	X	X	X	
SNF services		X	X		X	
Charges		X			X	X
Amounts paid		X	X	X	X	X
Dates of service		X	X	X	X	X
Rev. center codes		X	X	X	X	
DRG		X	X	X	X	
Provider codes		X	X	X	X	X

X\* county, state and zip

X\*\* requires special approval

**Table 2. Incidence Cancer Cases in Georgia, 1999-2005,  
by Data Source**

<b>Data Source</b>	<b>Total Enrolled Population (e.g., for 2004)</b>	<b>Breast Cancer</b>	<b>Colorectal Cancer</b>	<b>Lung Cancer</b>	<b>Non-Hodgkins Lymphoma</b>	<b>Prostate Cancer</b>
<b>GCCR (total GA incidence cases)</b>	<b>N/A</b>	<b>35,835</b>	<b>25,190</b>	<b>37,351</b>	<b>8,602</b>	<b>35,892</b>
<b>GCCR-Medicare</b>	<b>504,000</b>	<b>10,622</b>	<b>11,461</b>	<b>18,150</b>	<b>3,294</b>	<b>16,289</b>
<b>GCCR-Medicaid</b>	<b>362,390</b>	<b>7,137</b>	<b>2,235</b>	<b>1,486</b>	<b>455</b>	<b>1,565</b>
<b>SHBP</b>	<b>678,751</b>	<b>2,863</b>	<b>1,085</b>	<b>1,202</b>	<b>445</b>	<b>2,956</b>
<b>KPG</b>	<b>285,000</b>	<b>1,203</b>	<b>456</b>	<b>505</b>	<b>187</b>	<b>1,242</b>

**Table 3. Distribution of Coverage  
Across Health Plan Types for  
GA SHBP (2005)**

<b>Plan Type</b>	<b>0-65</b>	<b>66+</b>	<b>Total</b>
CDHP	573	9	582
HMO	282,826	3,882	286,708
Indemnity (FFS)	9,977	4,738	14,715
Other	95	485	580
PPO	280,292	50,693	330,985
Total	573,762	59,808	633,570

# *Next steps...and further down the road*

q Dive into the work before us – a tall order, and we're at the starting line.

*Then...*

q Inform, learn from, set the stage for *recruiting* the other major private *health plans* in Georgia

- Initial discussions have begun with another major payer in the state
- Ongoing contact with Georgia Association for Health Plans

q Build capacity to incorporate *biomarker data* into the bilateral linked data sets and the GA Consolidated Cancer Data Resource by capitalizing on

- Biorepository Alliance of Georgia for Oncology (BRAG-Onc) *{prospective tissue samples}*
- SEER Residual Tissue Repository (RTR) Program *{captures discarded tissue, linked to SEER data}*

## *....and further down the road*

n Collecting and linking *patient-reported outcomes* (e.g., health-related quality of life, symptom bother, satisfaction with care, economic burden) by sampling patients from GCCR or other frames

- Could be done now -- Prostate Cancer Outcomes Study (many others) show the way

- As with all other aspects of registry data augmentation, patient protection is paramount

q Reducing time lags between receipt of care, data reporting, feedback, and analysis – moving toward *rapid case ascertainment* by embracing leading-edge initiatives

- Georgia Cancer Quality Information Exchange

- Commission on Cancer's Rapid Quality Reporting System (RQRS)

*Enhanced RCA will also assist identification of clinical trial candidates (a major aim of Georgia Center for Oncology Research and Education (GA CORE))*

# *Prospects for Sustainability*

## **n Building state-level data infrastructure requires start-up funds...and if you build it, will they (the funders) eventually come?**

- Publicly and privately supported research on CER, CEA, QOC, population disparities in access and outcomes (non-experimental, observational research designs)

- Potential platform for practical clinical trials (experimental research designs)

- Empirical base for post-marketing regulatory studies, e.g., Phase IV

## **q In the meantime, one keeps the enterprise afloat through**

- Grants for research and data infrastructure development

  - (e.g., the ASPH/CDC award supporting the current data-linking work in GA)

- Collaborations that can provide some longer-term sustenance and stability

  - (e.g., in its new 7-year renewal application to NCI, the GA SEER program features “Augmenting Georgia Cancer Registry Data” as key initiative)

# *Project Team*

**Joseph Lipscomb, PhD, Emory**

*Principal Investigator*

**Kathleen Adams, PhD, Emory**

*Co-Principal Investigator*

**Cathy Bradley, PhD, Virginia Commonwealth University**

**Amy Chen, MD, Emory**

**Christopher Flowers, MD, Emory**

**Theresa Gillespie, PhD, Emory**

**David Howard, PhD, Emory**

**Douglas Roblin, PhD, Kaiser Permanente of Georgia**

**Kevin Ward, PhD, Emory (and Director of the Georgia SEER Program )**

## Contractors (institutional)

**Thomson Reuters Healthcare**

**Kaiser Permanente of Georgia**

## Data Development and Exchange Agreements

**Georgia Department of Community Health, including these components:**

**GCCR, Medicaid program, State Health Benefit Plan**