
Data Linkages to Improve Health Outcomes

An End User Perspective

Deborah Schrag MD
Department of Epidemiology and Biostatistics
Department of Medicine
Memorial Sloan-Kettering Cancer Center

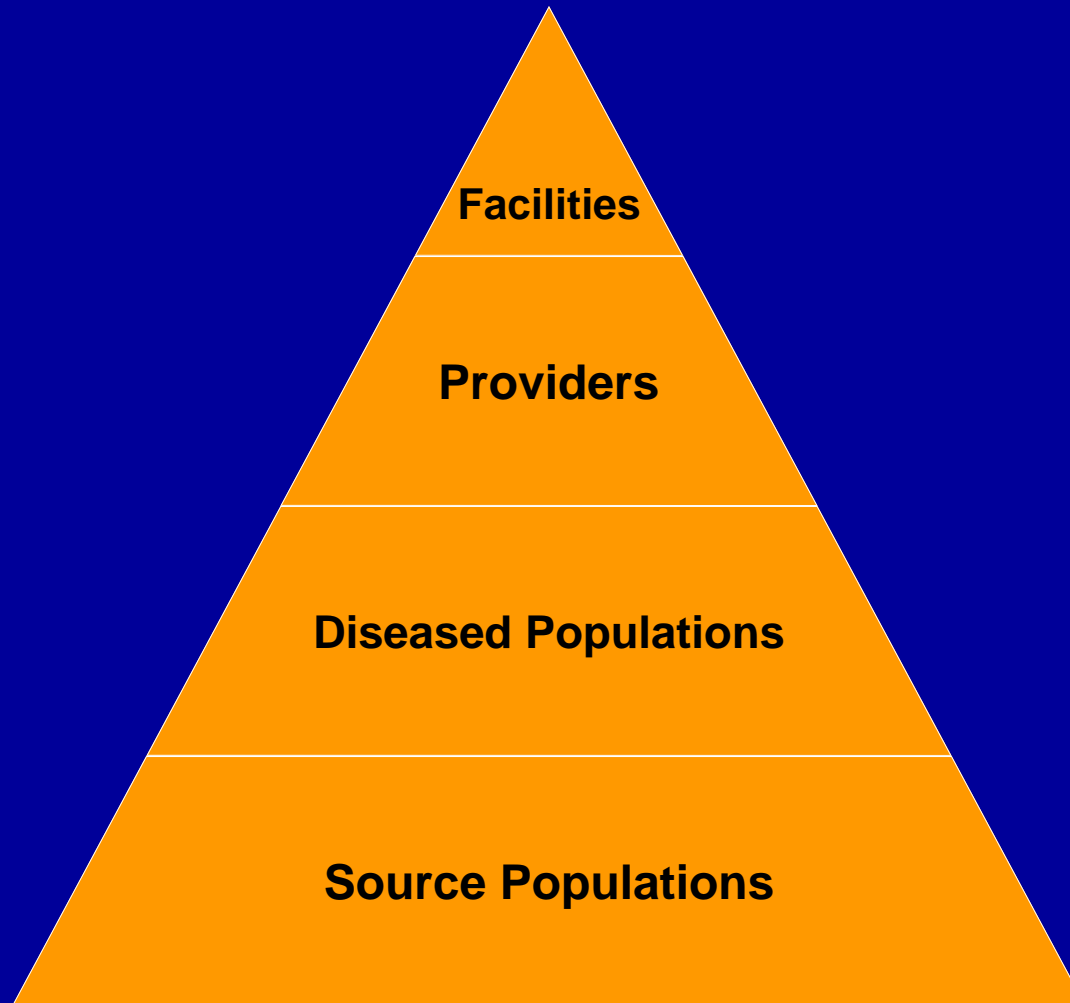
Overview

- **Types of research questions**
- **Examples of linkage attempts**
- **Challenges encountered**
- **Wish list**

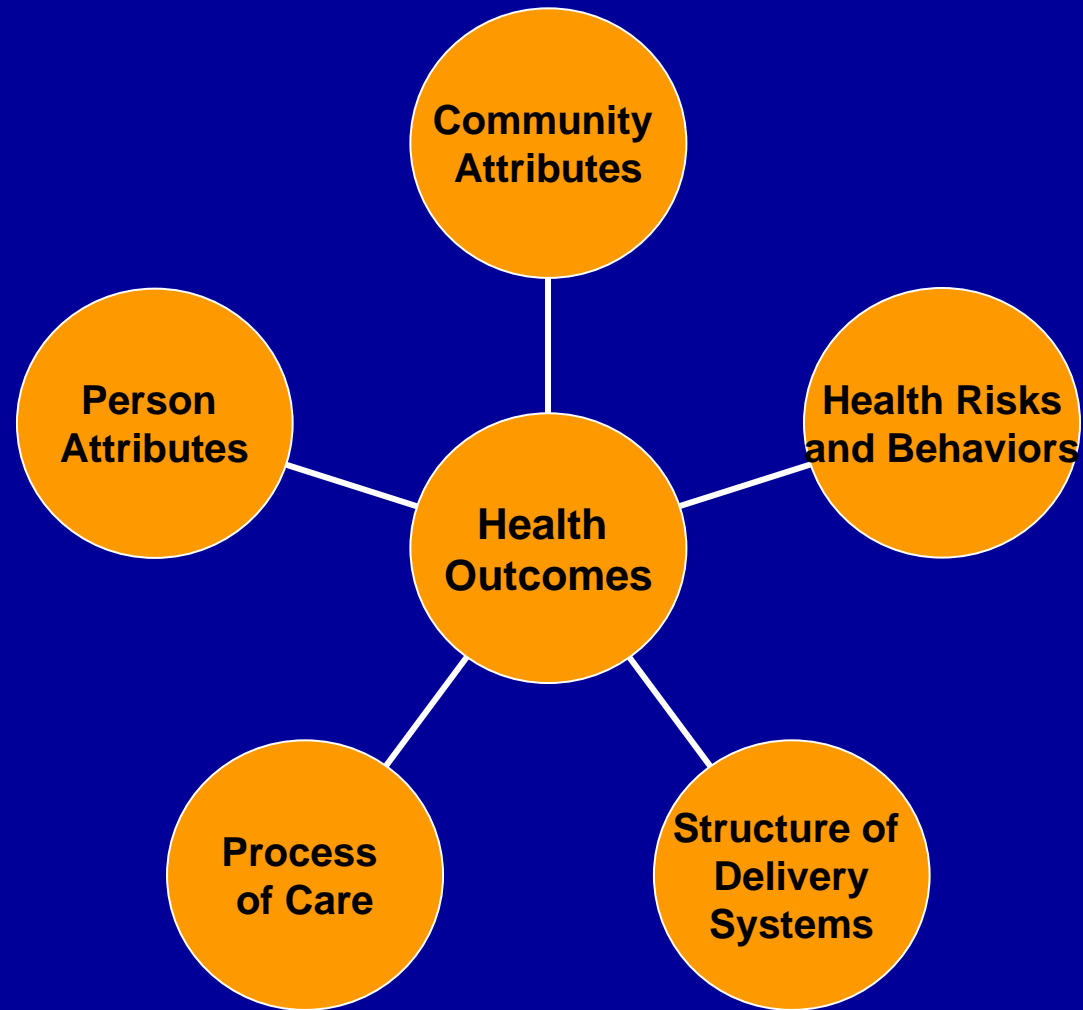
Health Services Research

- Examines relationships between need, demand, supply, delivery and outcomes of health care:
 - Disparities in care
 - Access/barriers to care
 - Technology dissemination
 - Quality measurement
 - Efficiency of care delivery

Layers of Data



Evaluating the Quality of Health Care



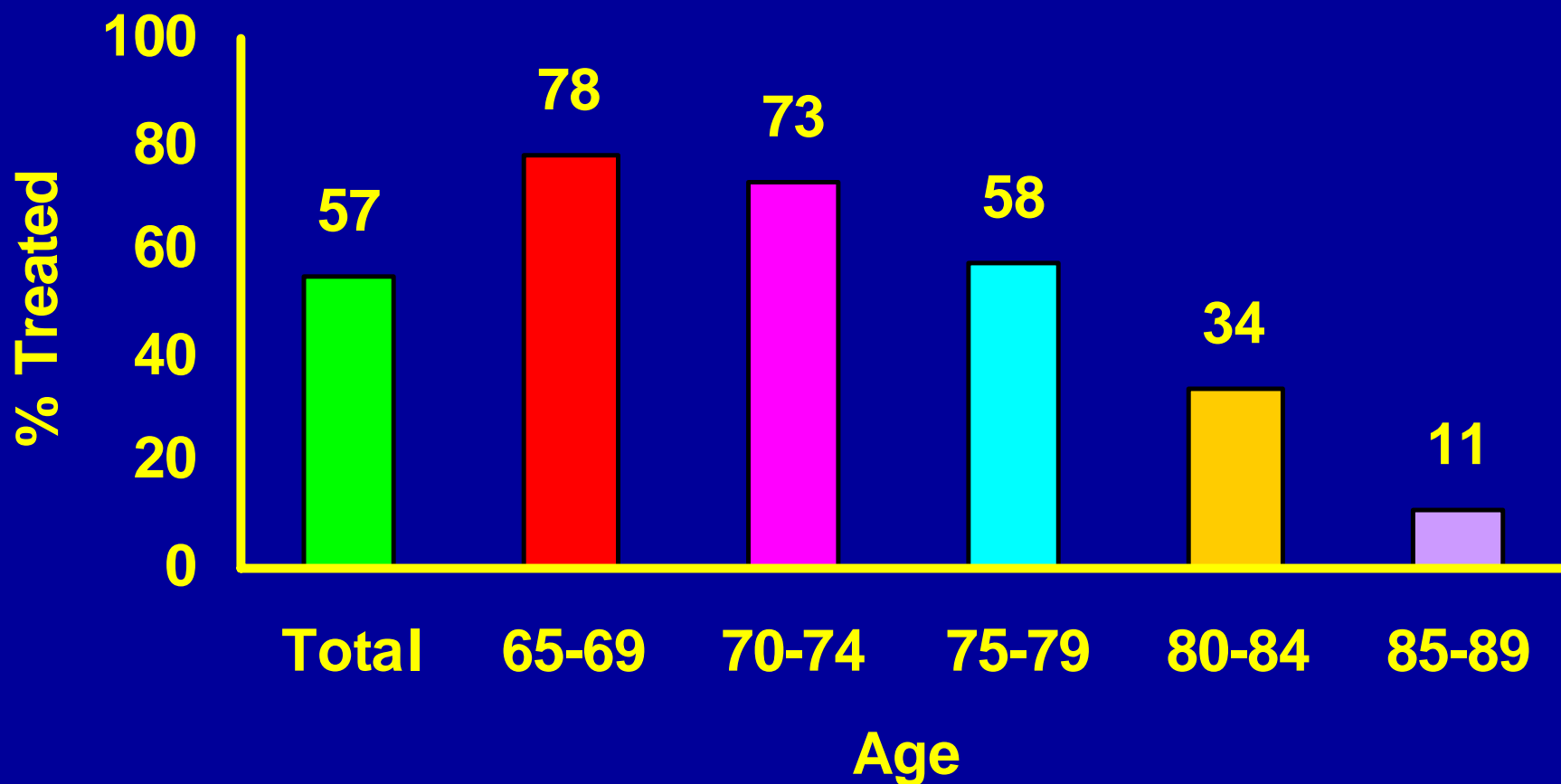
The Spectrum of Sources for Clinical Research Data

- **Population-Based data**
 - All patients in NY State with lung cancer
 - Data Source: Cancer Registry and Census Data
- **Quasi-population-based data**
 - All patients in NY State with lung cancer covered by Oxford
 - Data Source: Medical records and claims for care from Oxford
- **Non-population-based data**
 - All NY State patients evaluated at MSKCC with lung cancer
- **Health Services Research Strategy:**
 - Simultaneous use of various data sources and juxtaposition of analyses can reveal opportunities for improving health care delivery

Implementation Gap

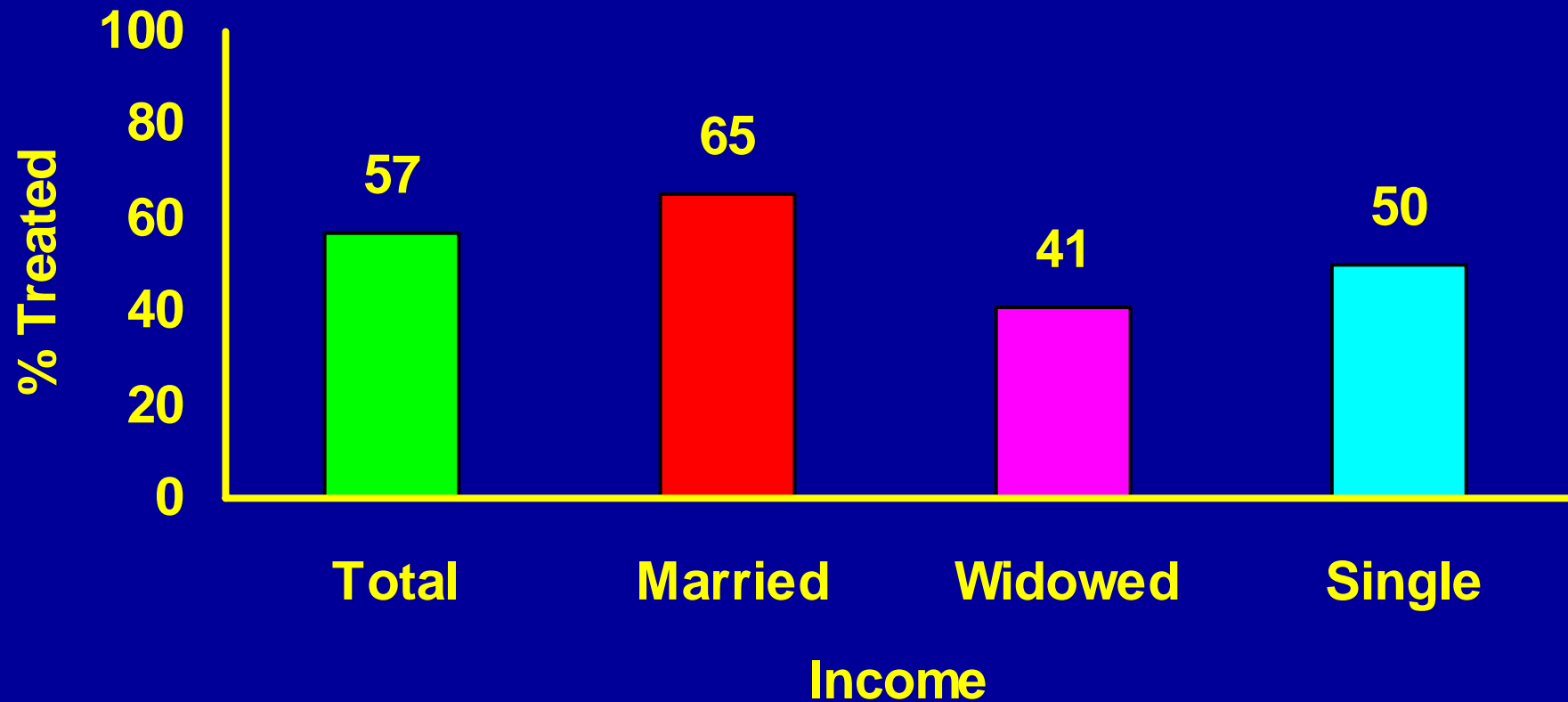
- **Efficacy – Effectiveness = Implementation Gap**
- **Need to understand reasons for gaps**
- **Identify important and remediable sources of variation**
 - **Endogenous to patients**
 - **Endogenous to MDs**
 - **Endogenous to health care system**

Adjuvant Chemotherapy Use By Age Stage III Colon Cancer in SEER-Medicare



P (adj) < .001

Stage III Adjuvant Chemotherapy By Marital Status



(p adj <0.001 HR for single vs.
married 0.62, 95% CI 0.54-0.72)

Linking to Providers

- **Why don't all patients get chemotherapy?**
- **Do they refuse?**
- **Do they see a medical oncologist post-operatively?**
- **UPINs on CMS claims can be linked to CMS specialty codes but the data are incomplete**

Medical Oncology Visits

Medicare Claims Indicate Provider Specialty



Most patients who forego adjuvant therapy make treatment decisions without oncology consultations. Provider specialty cannot be precisely specified----some oncologists are internists

Wish List: Provider Characteristics

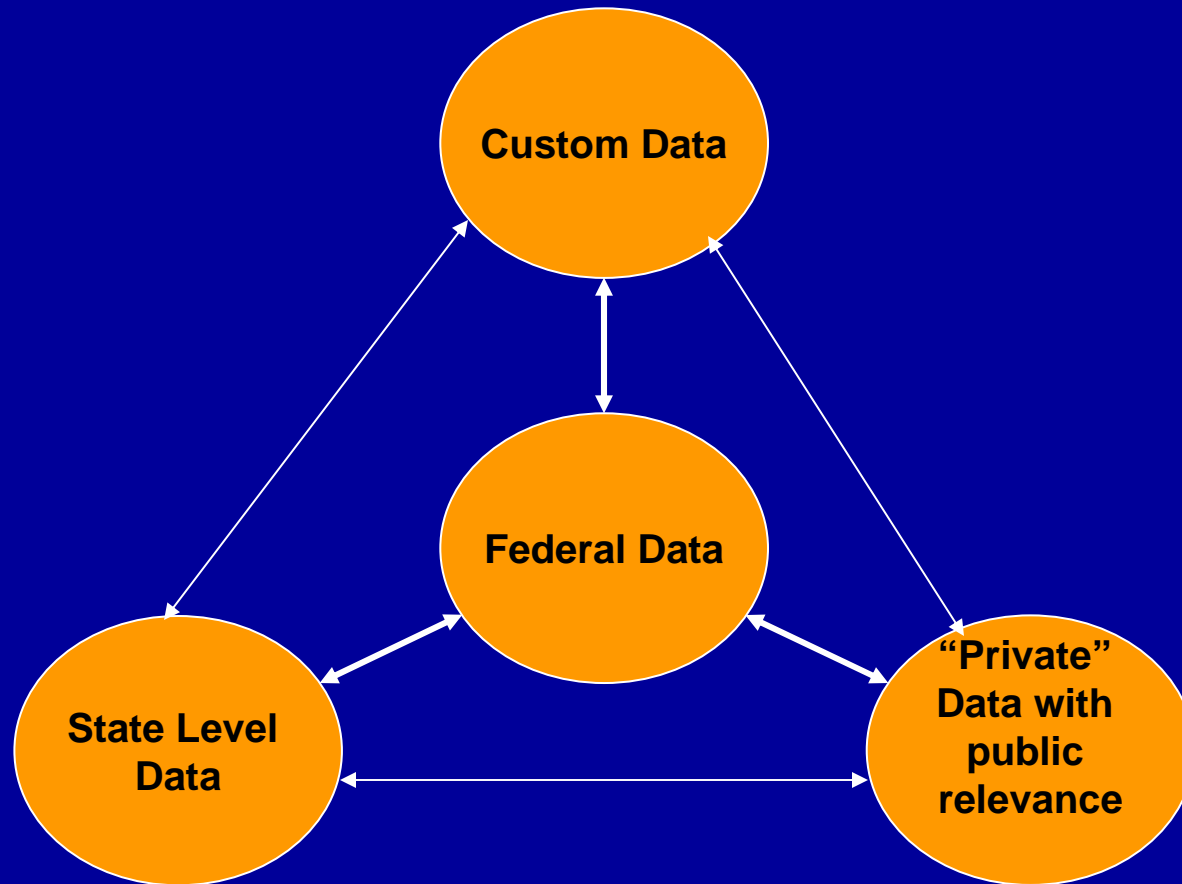
- **Linkage of UPINs on claims data to files describing physician characteristics**
- **AMA data is better than CMS data**
- **ABIM/ACS is better than AMA**
- **State-level data is most complete and most difficult to obtain**

Pharmacy Claims

- Oral chemotherapy?
- Anti nausea medications?
- Adherence to therapy
- Pain control

- Wish list:
 - Part D data
 - Medicaid data
 - Private claims data sets

Taxonomy of Data Custody Types



Capacity for Mammography in the US

- US women age 40-80 need mammograms
- Many women unscreened
- Large racial disparities
- Lack of facilities and radiologists are potential reasons for sub-optimal use

- Does lack of capacity explain geographic variation in use?
Racial disparities?

- Does capacity for mammography predict breast cancer incidence and mortality?

- Geocoding

Data Sources

- **Where are the facilities? FDA accreditation data**
- **Where are the radiologists? AMA/States**
- **Where are women unscreened? BRFSS, Medicare**
- **Where are the high rates of breast cancer? SEER**

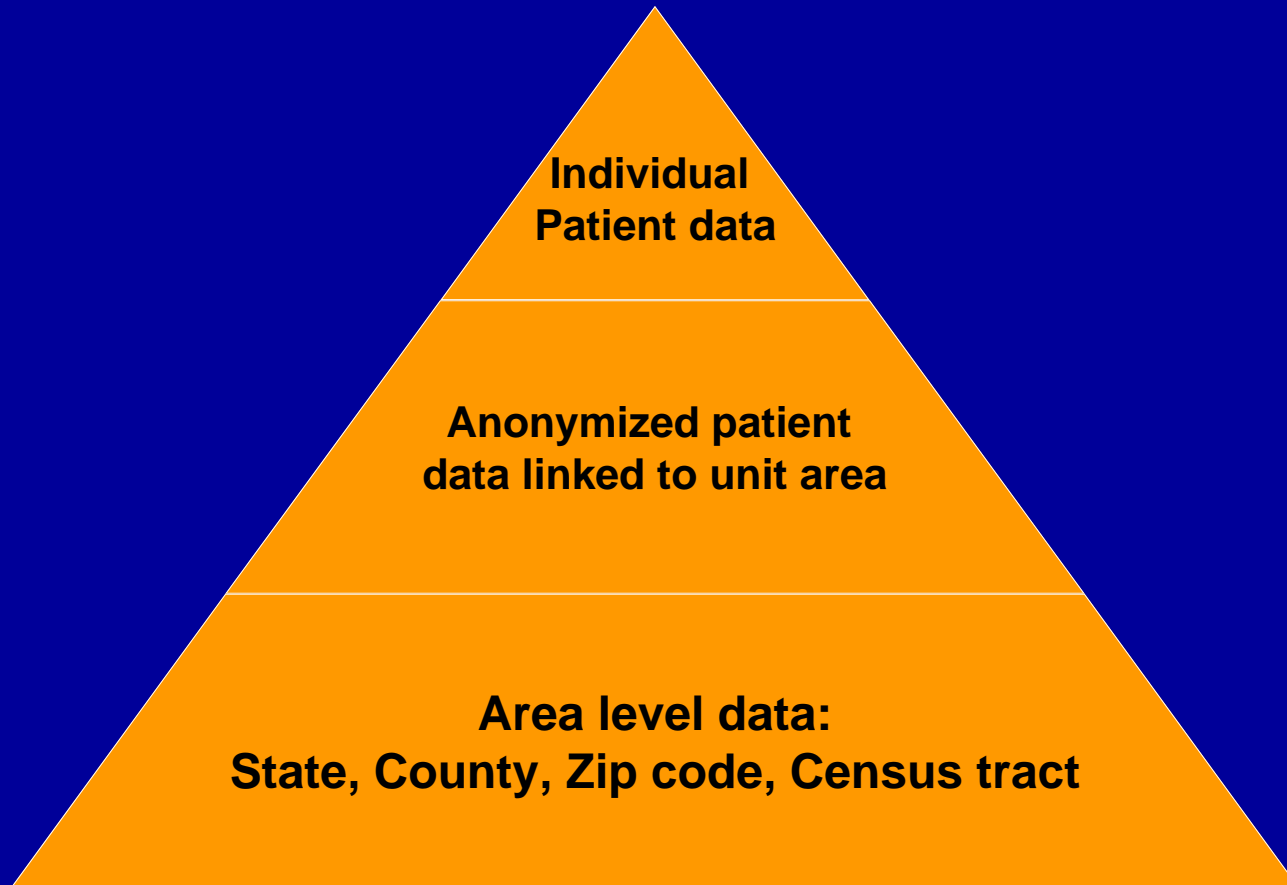
- **Data desired at census tract level**
- **Where to start to obtain permissions?**
- **Approval from one agency or many?**

- **Central clearinghouse, clearly delineated procedures would help**

Area versus Person-Level Data

- **Access to granular area level data helps most health services researchers**
- **Privacy/security concerns involve less risk**
- **Enables researchers to extract greater information from their own person-level data sets**

Layers of Data Access



Repetitive Common Tasks

- **Geographic variation:**
- **Where are? :**

- **Patients**
- **Providers/services**
- **Disparities**
- **Mortality rates**

Wish List

- **Access to choropleth maps**
 - **By county, zip code, census tract**
 - **Useful for common data elements census/survey data results**
 - **Shared resource for investigators**
 - **ArcGIS software**

Fragmentation of Care

- Do patients with chronic conditions in NY State where there are many hospitals consolidate their care or is it fragmented across multiple institutions?
- Is fragmentation higher in the Medicaid program?
- **SPARCS: statewide discharge database**
 - Available from states
 - Some states have data available from NCHS
- State discharge data not linked to patient residence/census data
- Medicaid enrollment data

Medicaid

- **Largest component of state budgets**
- **Health care for poorest, often sickest members of society**
- **Untapped resource because of complexity of data structure, organization and access, completeness**
- **Enrollment versus process**

The perfect as the enemy of the good

1998 Incident Primary Cancers Reported to CCR and % in Medi-Cal

Tumor Site	Incident Cancer Cases from the California Cancer Registry			Cases in CCR and Medi-Cal	
	All 1998 Cases	% age 18-64 at Diagnosis	1998 Cases, Age 18-64	Medi-Cal and CCR	As % of All Cases, Age 18-64
Breast	20,864	54.6%	11392	1014	8.9%
Lung	17,004	30.6%	5203	997	19.2%
Colorectal	10,254	27.3%	2796	557	19.9%
Cervical	1,690	79.9%	1350	287	21.3%
Prostate	19,001	29.5%	5605	268	4.8%
Testis	937	93.6%	877	78	8.9%
Uterine	3,587	48%	1722	176	10.2%
Bladder	5,452	26.5%	1445	113	7.8%
Hepatoma	1,570	53.3%	836	288	34.4%
Stomach	2,552	32.2%	822	197	24.0%

Duration and Timing of Medi-Cal Enrollment in Relation to Cancer Diagnosis

Number of patients recorded in CCR with cancer diagnosis between 7/97-6/98 with records in 1/97-12/98 Medicaid files	# of patients 7264	% of cohort 100%
Duration of Medicaid Enrollment over 24-month interval		
Entire 24 months	3443	47%
13-23 months	1671	23%
7-12 months	1540	22%
1-6 months	610	8%
Medicaid Enrollment Status at month of diagnosis		
YES, Enrolled during month of diagnosis	5364	74%
NO, First enrolled after month of diagnosis*	1698	23%
NO, Enrolled prior to diagnosis but not during month of diagnosis	202	3%

*Of these, 905/1698=53% enrolled within 3 months of diagnosis.

How often do California Medicaid claims corroborate cancer site specific diagnoses reported to the CCR?

Cancer Type	# of CCR-Medicaid enrollees with 1998 cancer diagnoses	# (%) of Patients diagnosed at anytime during 1998 with a corroborating diagnosis code recorded in 1998 Medicaid claims files	# of Patients Enrolled in Medicaid in 1998 with CCR cancer diagnosis in the <i>first 6 months</i> of 1998	# (%) of Patients from preceding column who have a cancer diagnosis code recorded in 1998 (entire year) Medicaid claims files
Breast	1014	770 (76%)	549	436 (80%)
Lung	997	751 (75%)	523	416 (80%)
Colorectal	557	394 (71%)	274	207 (76%)
Cervical	287	207 (72%)	146	117 (80%)
Prostate	268	176 (66%)	127	93 (73%)
Testis	78	60 (77%)	38	28 (74%)
Uterine	176	124 (70%)	86	64 (74%)
Bladder	113	86 (76%)	53	43 (81%)
Liver	288	180 (63%)	151	99 (66%)

Medicaid Data

- **SEER-Medicaid Data**
 - Attempted link in California
 - 2 years to obtain data sets
 - Denominator file structure limits ability to identify cohorts of the chronically poor
- **Challenges:**
 - Retroactive enrollment
 - Chronic vs. episodic poverty
 - Spend downs—illness precipitates enrollment
 - Variation in states thresholds/generosity
 - Definition of an HMO

Wish List: Medicaid Data

- **Consistent definitions in Medicaid enrollment files**
 - What does managed care mean?
 - When are claims itemized?
- **Linkages of Medicaid data files to state discharge abstracts**
- **Geocoding of where Medicaid beneficiaries reside**
- **Linkage to pharmacy data**
- **Linkage to census tract socioeconomic variables**

Priorities

- **Coordination of procedures for obtaining access to data and the review process**
- **Standardization of reporting rules (e.g. N must not be less than 10)**
- **Develop categorization schema for types of linkages**
- **Central clearinghouse/index describing linkages that exist as well as those that are possible**
- **Facilitate federation of state data**
- **Chloropleth maps for use in commons based systems**
- **Work with states to facilitate analyses of Medicaid enrollment and claims files**