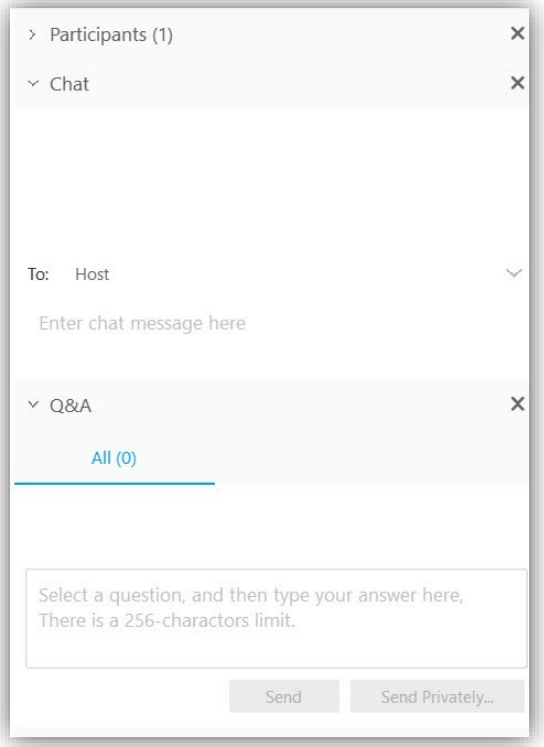
Using WebEx and Webinar Training Logistics



- All lines will be in listen-only mode
- Submit questions at any time during the presentation. Type into the Q&A Panel and select Host
- A moderator will ask the question on your behalf during the Q&A portions of the training
- This training webinar is being recorded and will be posted at a later date

Welcome to the Virtual SEER-Medicare Training Workshop

SEER-Medicare Training 2019

Beth Virnig, PhD, MPH

About this Course

- Goal: to provide a **high-level** overview of the linked SEER-Medicare data so that users will be able to:
  - 1. Understand how the SEER and Medicare data are combined, their structure, and potential uses
  - 2. Understand how to identify a study population, identify baseline characteristics, and use common inclusion/exclusion criteria
  - 3. Process for obtaining and publishing data

Note: This could easily be a 2-day, 3-day, week long course if we covered full details on all the topics included. We will focus on high-level concepts and will point you to resources for specific details…

vs.

Instructors all from the University of Minnesota

- Beth Virnig, PhD, MPH

- Stephanie Jarosek, RN, PhD

- Helen Parsons, PhD, MPH

- Kristi Swanson, MS

Agenda

- Morning Session
  - Overview of SEER-Medicare data
  - Defining your study population
  - Baseline measures
  - Measuring treatment
- Afternoon session
  - Measuring outcomes
  - Getting the data and publishing
  - Final advice

There will be opportunities for Q&A after each session

Overview of the SEER-Medicare Data

Segment 1

SEER-Medicare Training 2019
Beth Virnig, PhD, MPH

Opening Remarks

- Data from the Surveillance, Epidemiology and End Results (SEER) cancer registries are combined with data from the Medicare program to create the SEER-Medicare data
- Understanding the two programs is essential to properly interpreting the linked data and for identifying opportunities for effective use

SCHOOL OF PUBLIC HEALTH
UNIVERSITY OF MINNESOTA

Part 1: The SEER Program and Data

What is the SEER Program?

- The Surveillance, Epidemiology and End Results (SEER) Program operates registries that collect and report information about cancer incidence and survival in defined geographic areas (population-based)

- The National Cancer Institute (NCI) funds the registries which are the US' only comprehensive source of population-based cancer data that includes stage of cancer at diagnosis and patient survival

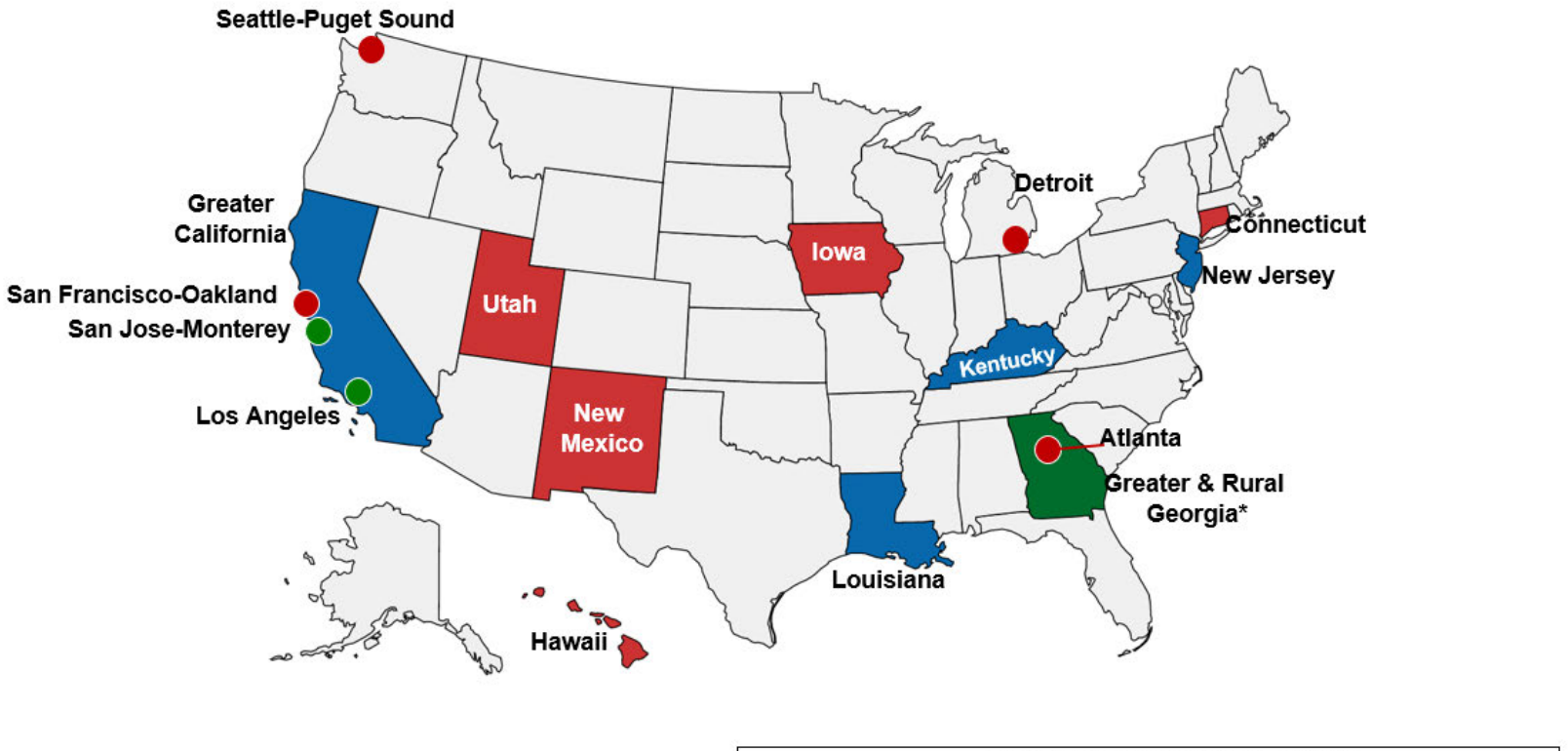
- Data sent to NCI are in a standard format. The data are essential sources of information for understanding cancer trends

Which Registries are Included in the SEER Program?

The geographic areas included in the SEER program have changed over time

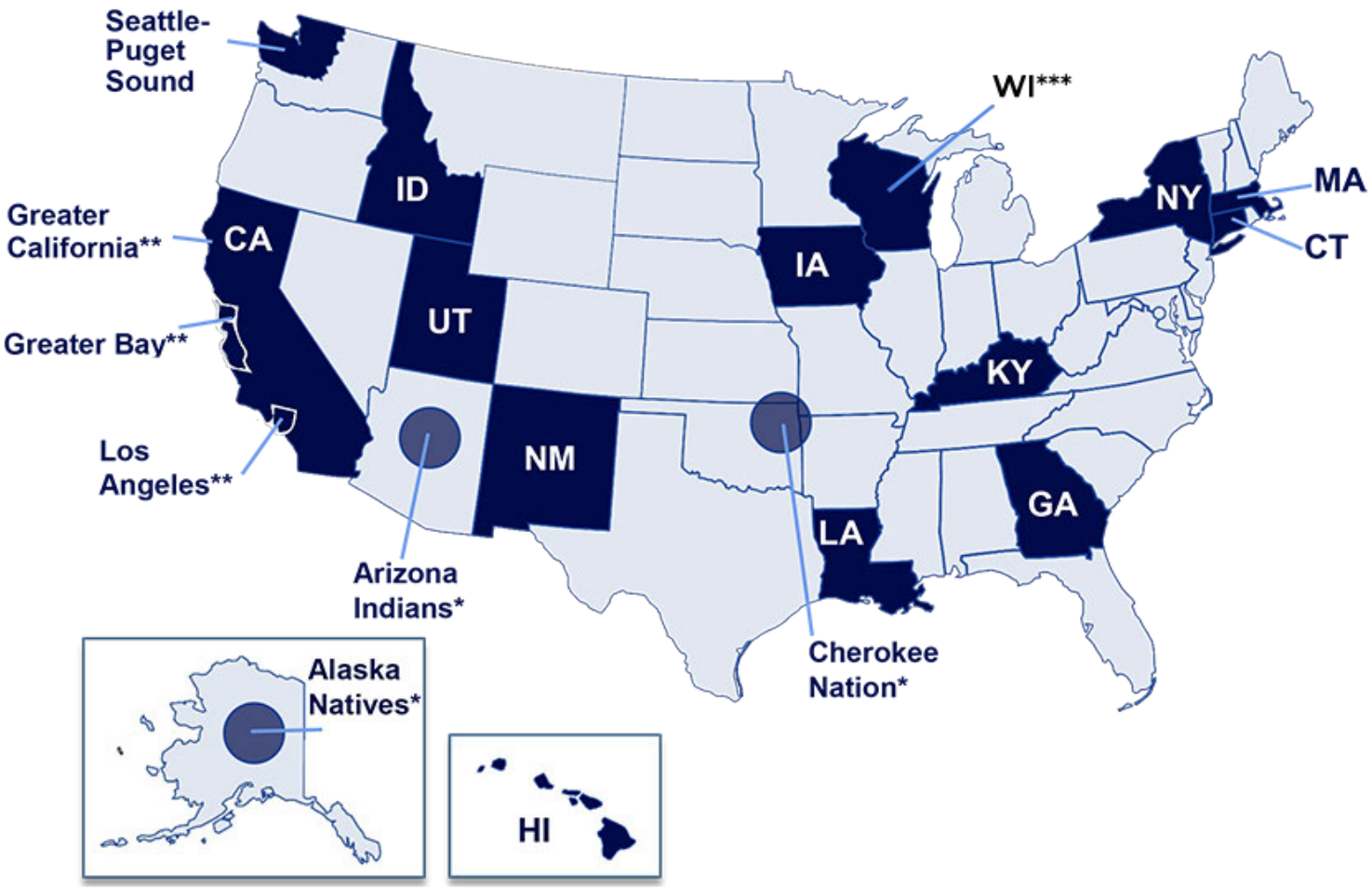
- 9 registries were part of the SEER program in the 1970's - 10% of the U.S. population covered
- Subsequent expansions
  - Current data available– 26% of U.S. population covered
  - Newest expansion (2018) will add 3 more states and greater representation of American Indians
- SEER registries are selected on the basis of ability to conduct high-quality cancer surveillance and to contribute to population representativeness or to include populations of special interest

SEER Registries 2001-2018



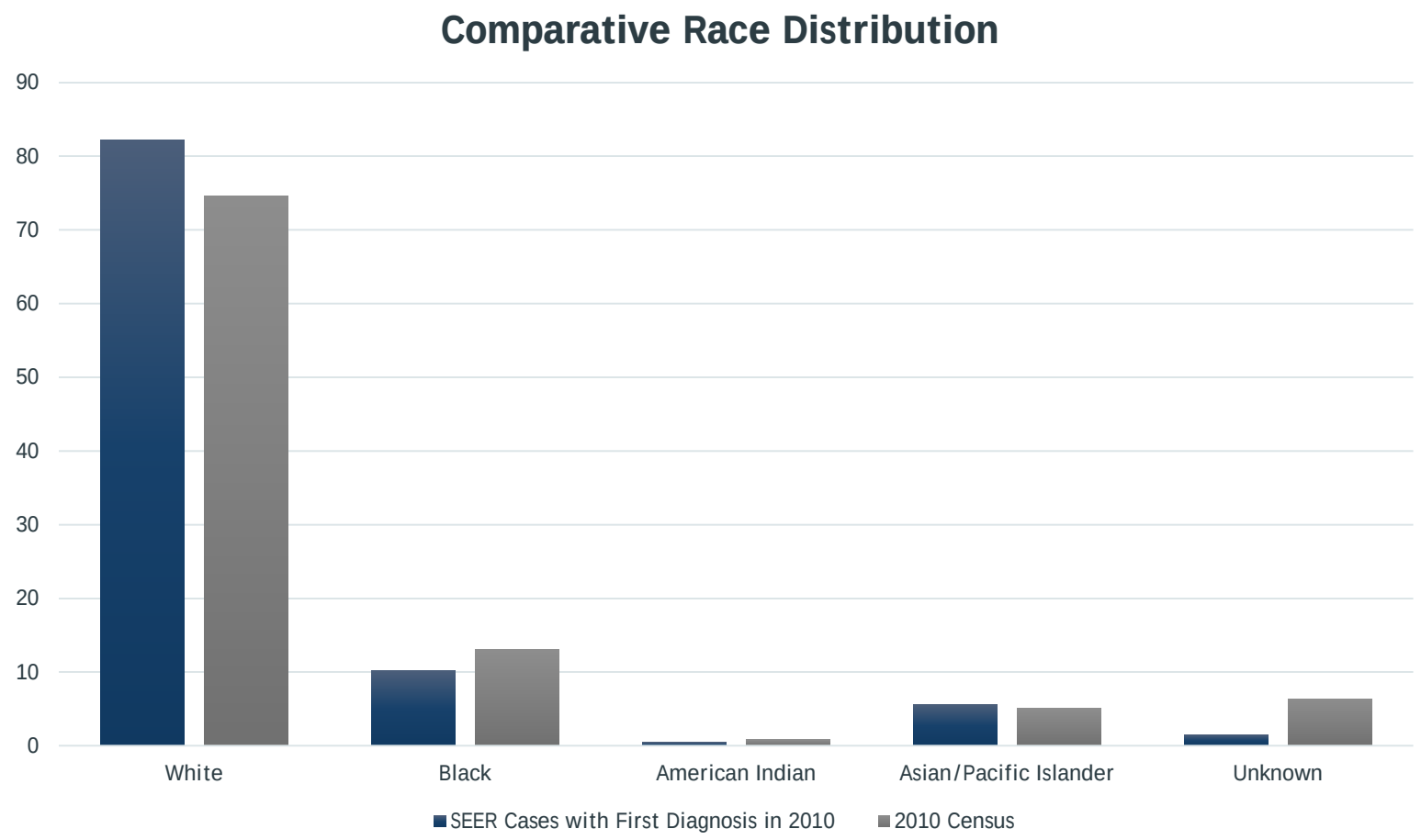
*Rural GA, Los Angeles and San Jose-Monterey data available since 1992. Greater GA data available since 2000.

SEER Registries (2018-present)



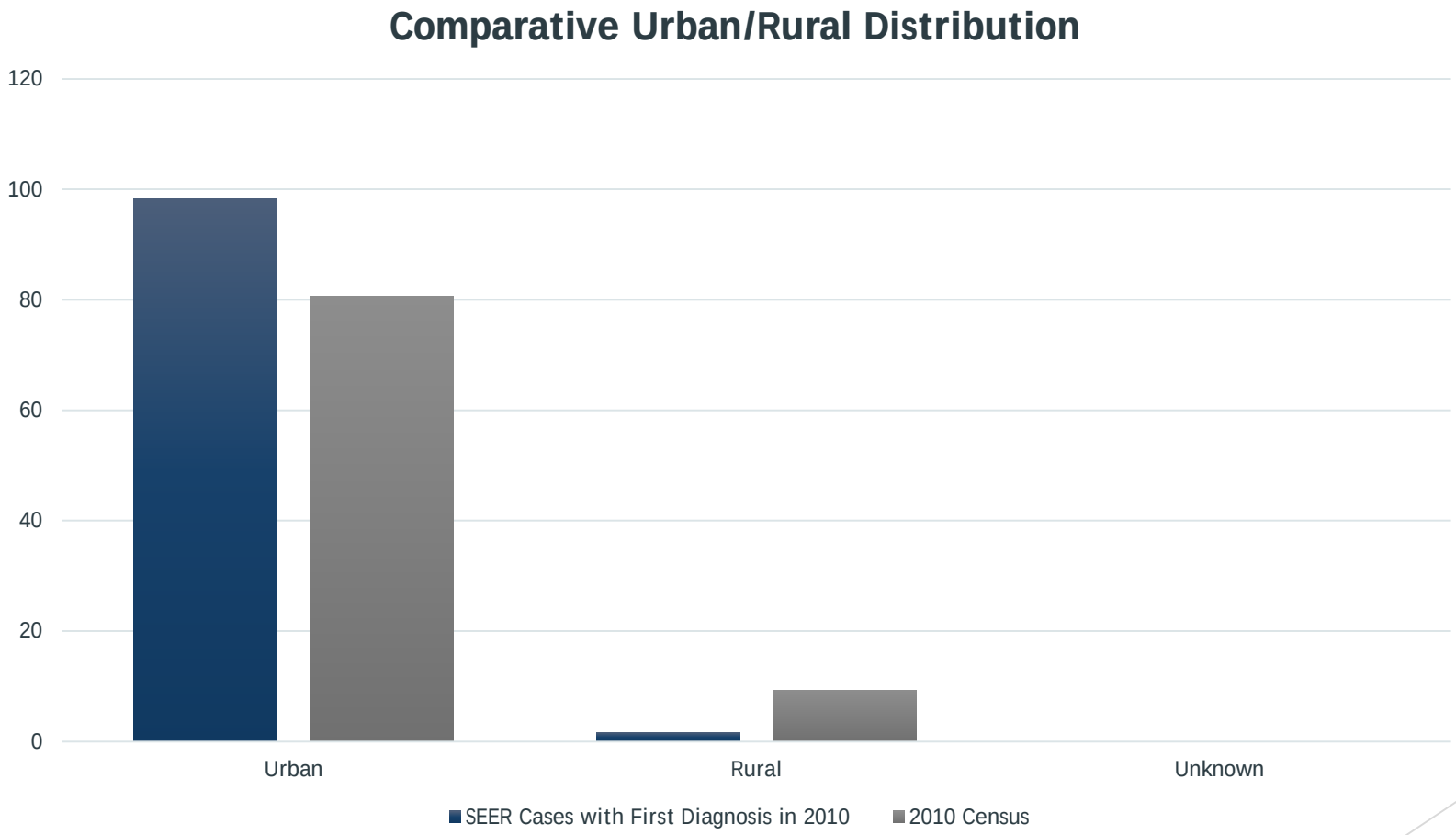
* Subcontract under New Mexico
** Three regions represent the state of California: Greater Bay, Los Angeles, and Greater California
***Research support registry only; not under contract to submit data

SEER Areas Represent a Diverse Population



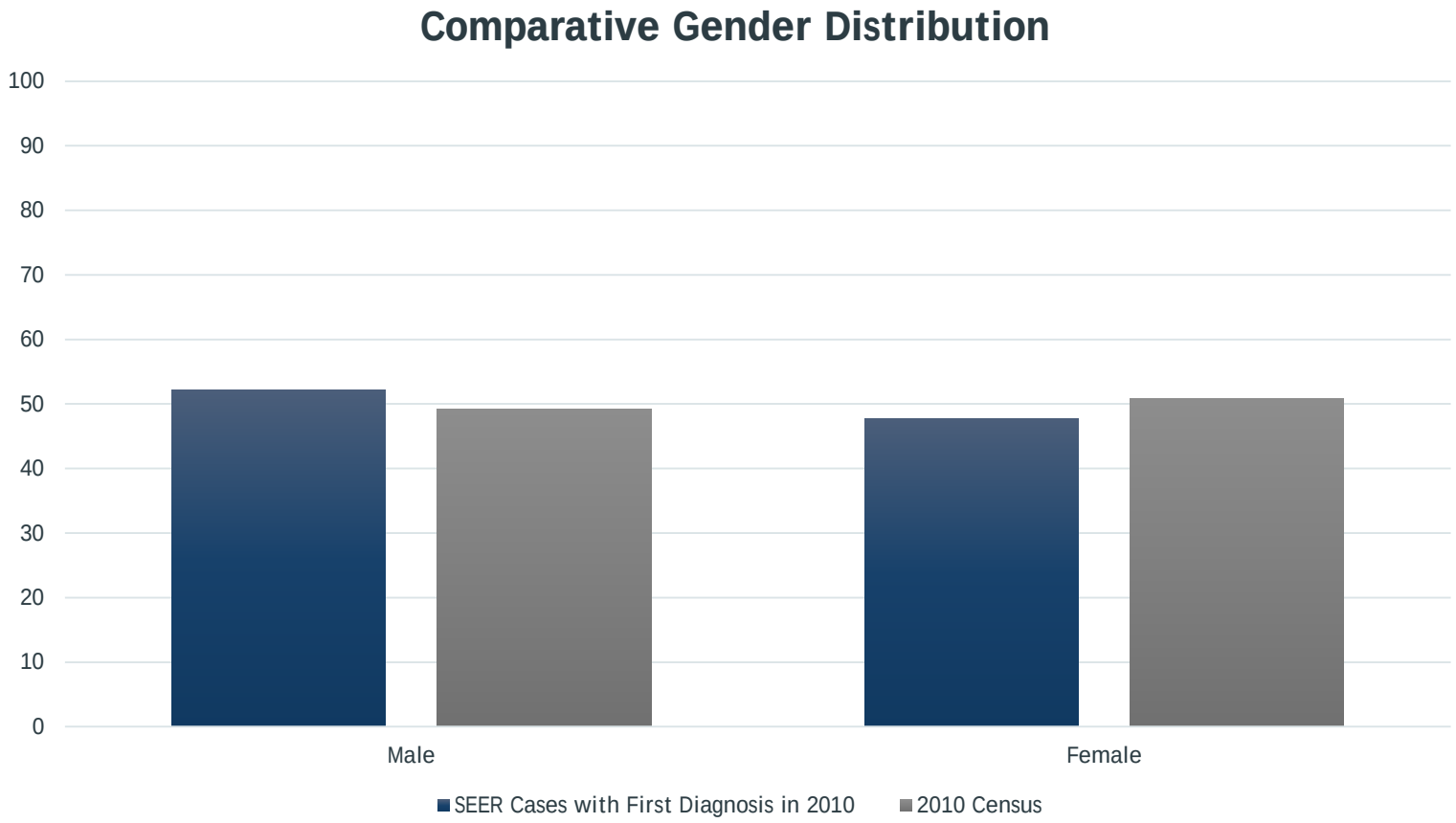
SCHOOL OF PUBLIC HEALTH
UNIVERSITY OF MINNESOTA

SEER Areas are More Likely to Represent Urban Locations



SCHOOL OF PUBLIC HEALTH
UNIVERSITY OF MINNESOTA

SEER Population Equally Represents Males and Females



How Cases are Identified

Registries use active and passive case finding methods

- Notification from facilities required by law
- Active review of source documents, such as pathology reports

The SEER program is primarily facility (i.e., hospital) based. Treatment provided only in physicians' offices may be less complete

Case Abstraction

- Data abstracted by trained professionals working for the hospital or the registry
- Consistent process involving agreed upon, established rules
  - As a user of SEER data, understanding rules is important to properly interpret data and to recognize opportunities

Data Collected by the SEER Registries

- 4 categories
  - Demographic
  - Information about the cancer
    - Identification of the cancer
    - Tumor related information
  - Treatment
  - Mortality

Demographic Information Collected by the SEER Registries

- Residence at diagnosis
- Sex
- Age at diagnosis
- Year of birth
- Place of Birth
- Race / Hispanic Origin
- Marital Status

Information About the Cancer: Identification of the Cancer

When and how was it diagnosed?

Date of Diagnosis (month/year)

Method of Diagnostic Confirmation

Who reported it to SEER?

Reporting Source

Was it their first cancer?

Sequence number (order for primaries for people with multiple cancers, 00=only 1 cancer)

Information About the Cancer: Tumor Related Information

- Cancer type
- Stage
- Histology
- Some tumor markers
- Lymph node testing and positivity
- Grade
- Size
- Extent of disease

Basic Information About First Course of Treatment

- First course of treatment:
  - All methods of treatment recorded in the treatment plan and administered to the patient before disease progression or recurrence.
- SEER collects more information than is released as part of the SEER-Medicare linkage
- Surgery (categories)
- Radiation therapy (categories)

- We will cover this more in segment 4…

What DOESN'T SEER Collect (or doesn't routinely release)

DETAILED INFORMATION ABOUT ALL TREATMENTS

INFORMATION ABOUT CHEMOTHERAPY

INFORMATION ABOUT TREATMENT BEYOND THE FIRST COURSE OF TREATMENT

CANCER RECURRENCE INFORMATION

EXACT TIMING OF TREATMENT

Everything You Ever Wanted to Know About SEER

- SEER website http://SEER.cancer.gov/
- Documentation for SEER Data https://seer.cancer.gov/data-software/documentation/
- Registrar training info http://training.SEER.cancer.gov
- SEER program and coding manuals
  - http://SEER.cancer.gov/tools/codingmanuals/index.html
  - http://SEER.cancer.gov/tools/codingmanuals/historical.html
- Comparative staging manual http://SEER.cancer.gov/manuals/historic/comp_stage1.1.pdf

Part 2: The Medicare Program and Data

What is the Medicare Program?

- The Medicare program is a federal health insurance plan available to qualifying elders (age 65 and older) and select disabled adults

Medicare Coverage

As an entitlement for most elders:

- Hospital Care, limited SNF, Hospice
- Called “Part A”
- 98% of elders have Part A

Optional coverage (requires premiums):

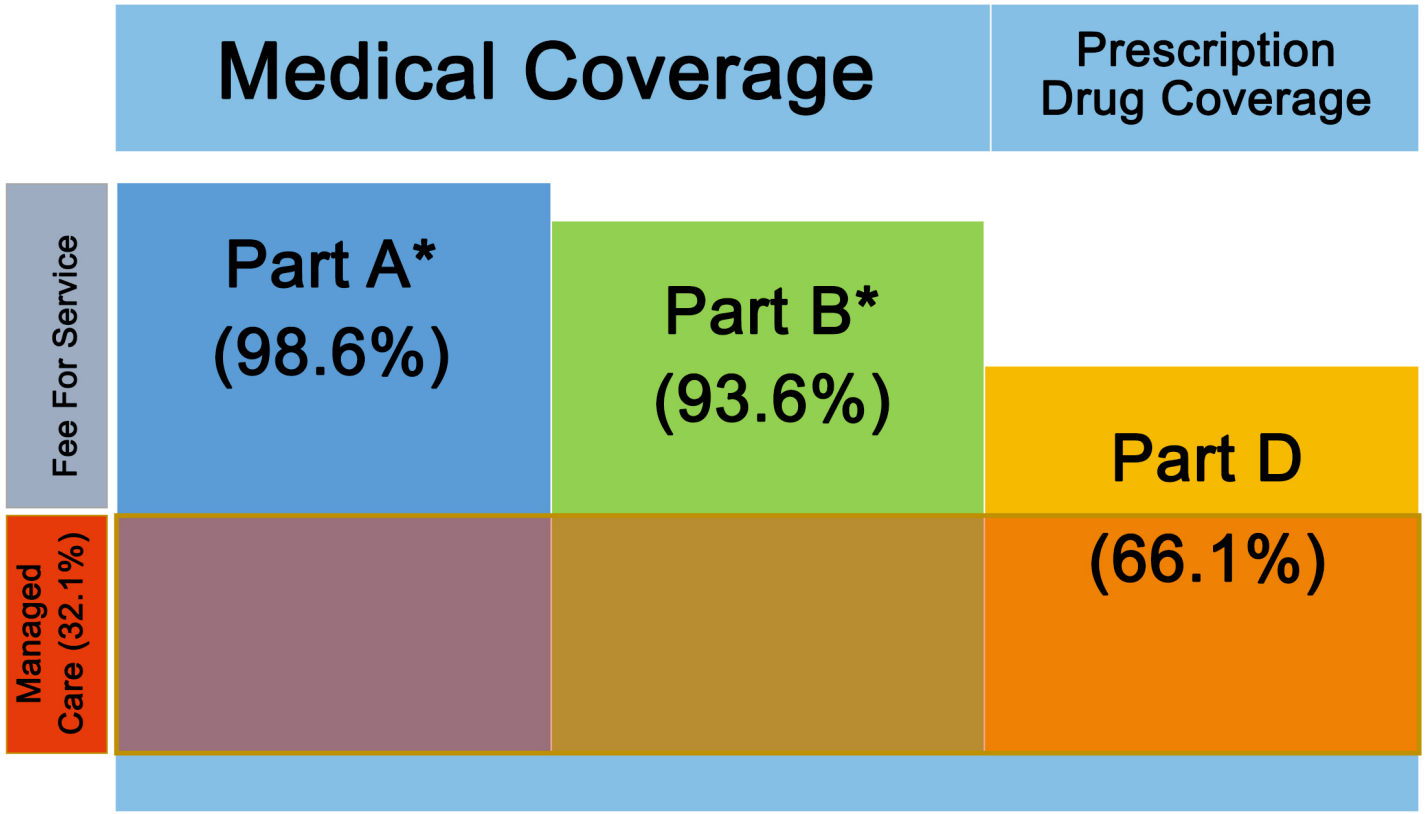
- Outpatient Care, Physician and Provider Care, Home Health Care
  - Called “Part B”
- 94% of elders enroll in Part B

Optional coverage (requires premiums):

- Prescription Medications
- Called “Part D”.
- Began in 2006, SEER-Medicare releases 2007+

Most People Over 65 Have Part A and Part B Coverage. Fewer have Part D or Have Their Care Coordinated Through Part C (managed care)

SEER cases not diagnosed by autopsy or death certificate, age 65 and older and alive as of Jan 1, 2014

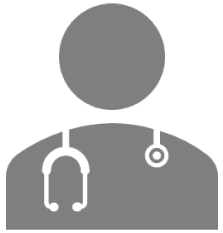


*92.8% covered under both Part A and Part B

What Are the Medicare Data?

- Medicare data contain two types of information:
  - Enrollment data
  - Data derived from bills submitted by providers and processed by Medicare

How Does That Work Anyway?



Whenever a Medicare enrollee receives health care, a bill is sent to the Medicare program to be paid.

The bill contains information about:

* who provided the care (a facility, MD, etc.)

*When the care was provided (date(s))

*why the care was provided (diagnosis)

*what care was provided (procedure)

*it also contains information about how much was paid for the care (covered in segment 5)

Organization of Medicare Data

- Billing/use data are divided across 7 files based on whether it is a Part A, B or D service and the billing form used
- 7 Claims-based services files
  - MedPAR (hospital and SNF)
  - Outpatient (non-inpatient facility)
  - NCH (also known as Carrier)
  - HHA (home health care)
  - Hospice
  - DME (durable medical equipment)
  - Part D (pharmacy)

Money in Medicare

- The fields in the Medicare claims files that address money are divided into two groups:
  - Charges
    - What the hospital, clinic, physician, etc. ASKED to be paid
  - Payments
    - What they were paid or approved to be paid

Details About Medicare Data That Make Them Particularly Useful

Contain all claims regardless of residence

The Medicare program collects all claims for fee-for-service care provided from the time a person enrolls in Medicare until death

Medicare claims can include care prior to the diagnosis, peri-diagnosis, and following diagnosis–depending when the diagnosis is made relative to Medicare enrollment date

But…

- Test reasons and results are not included:
  - Was the testing in response to some clinical sign or symptom?
- Test results not included:
  - PSA
  - Pathology
    - histology, margins,
- Behavioral risk factors rarely coded (except when the code is needed to justify care)
  - Obesity
  - Smoking
  - Family history
  - Alcohol use
- Conditions that are under-diagnosed in clinical settings will be under-reported in claims
  - Dementia, osteoporosis

What Information Could You Find in the Medicare Data?

Real world example – an elderly woman with breast cancer

Pre-diagnosis medical conditions that might influence treatment options (i.e., comorbidities)	Diagnostic service- mammography, u/s, CT/MRI, bone scan	Surgery related- type of procedure, surgeon, length of anesthesia time, blood use, length of hospital stay, secondary surgery, complications such as major infection
Other treatment- RT, chemotherapy	Provider info- hospital characteristics (teaching status, bed size, etc.), volume	Ambulance services
Consults	Rehospitalization	Rehabilitation
Other health care that occurs at the same time—cardiac stress test, diabetes monitoring, etc.	Treatment that she receives following completion of her cancer treatment	Payments for the above services

Structure of the Medicare Data: One to Many

PEDSF	
Patient A	Age, gender, Cancer(i), stage(i), etc.
…..	…….

MedPAR	
Patient A	Hospitalization 1
Patient A	Hospitalization 2
Patient A	…….

Outpatient (or similar – e.g., NCH/Carrier)		
Patient A	Claim 1	Service 1
Patient A	Claim 1	Service 2
Patient A	…..	…..

Bottom Line

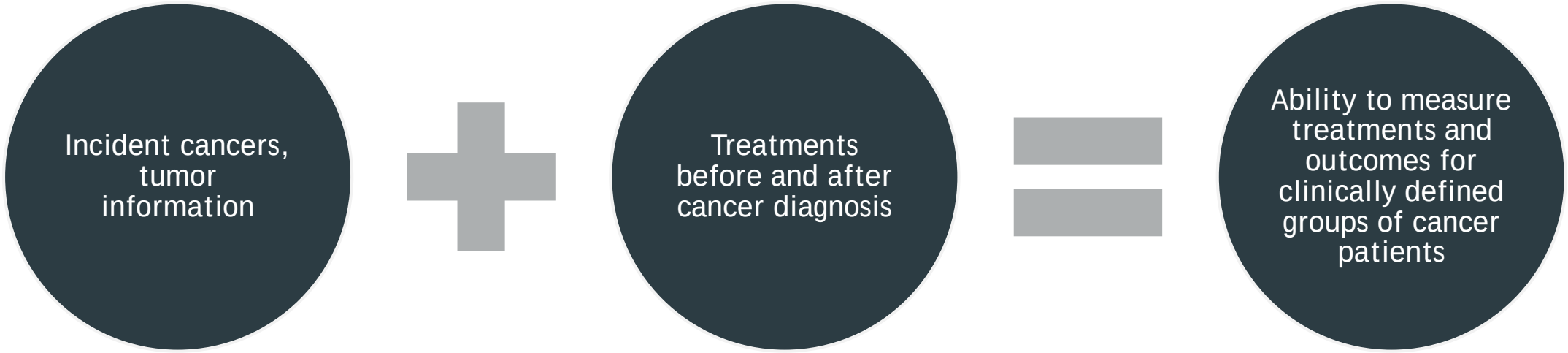
Medicare data are a rich source of information about cancer care and outcomes in the elderly.

However, they are complex data and it is important to understand what services are covered and how they are coded in order to properly interpret what you find.

Resources

- “Medicare & You” is a great resource for information about the Medicare program enrollment options and benefits. The 2019 version can be found here:
- https://www.medicare.gov/sites/default/files/2018-09/10050-medicare-and-you.pdf
- You can also search the internet for other years to identify benefit information specific to your study period.
- ResDAC.org

The SEER-Medicare Linked Data


Incident cancers, tumor information
Treatments before and after cancer diagnosis
Ability to measure treatments and outcomes for clinically defined groups of cancer patients

Linking the SEER-Medicare Data

Personal identifiers for people in the SEER registries are matched against CMS' master enrollment file to find their unique Medicare number

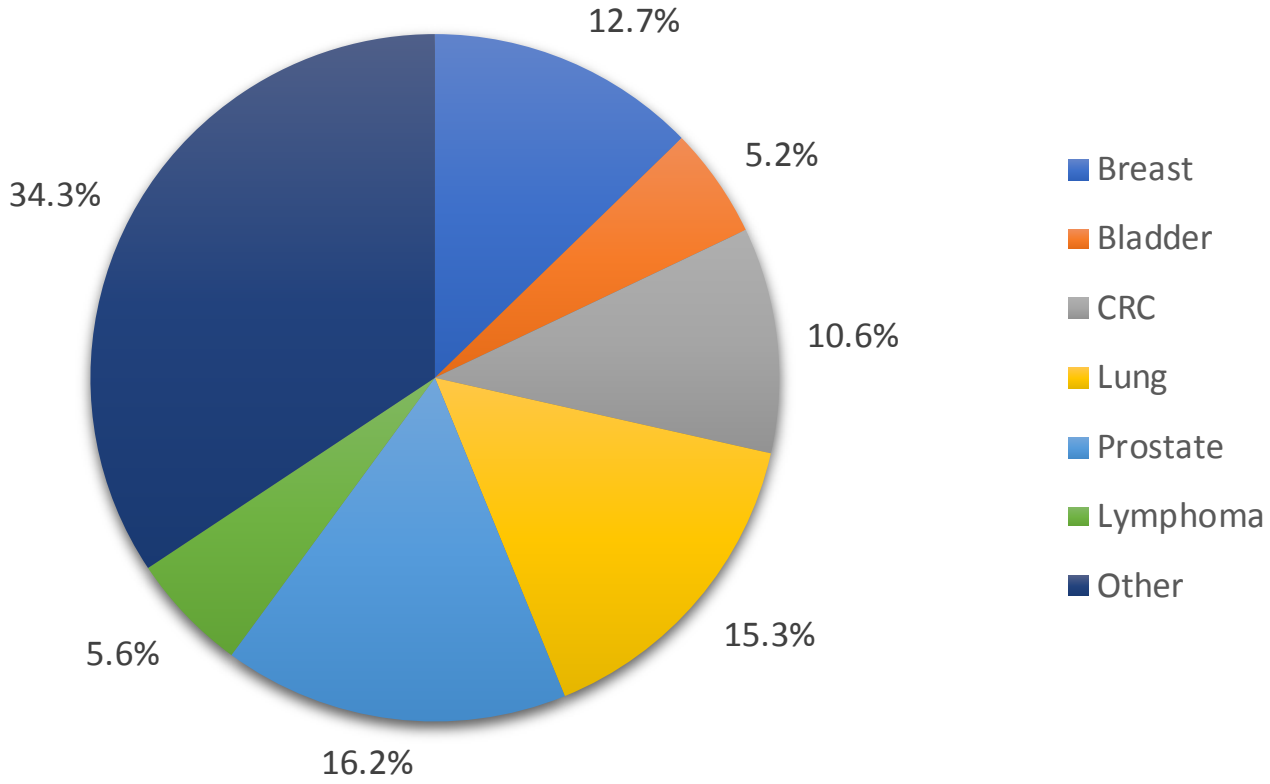
The unique Medicare number (HIC) is used to extract Medicare claims for the persons in SEER found to be Medicare beneficiaries

CMS sends files to IMS, NCI's programming contractor, for creation of analytic files. IMS removes the unique Medicare number and overwrites it with each individual's SEER case number

Linkage

- The registries send NCI all of their cases for the linkage
- Match rates for persons 65+ in SEER to Medicare for cases included in the most recent linkage: 95.7%
- Variation across registries ranging from 91.1% to 99.2%

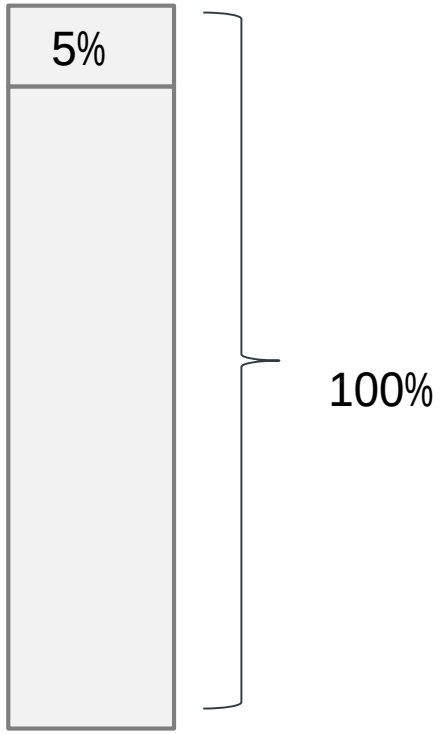
Incident Fee-for-service Cases Age 65+ in the SEER-Medicare Data by Cancer Site, 2004-2013 (N=1.1 Million)



Persons and Files Included in the SEER-Medicare Data

Persons Included in the SEER-Medicare Data

- 100% of patients in the SEER data who are found to be Medicare eligible
- A sample of Medicare beneficiaries residing in SEER areas who have not been reported as a cancer case to the SEER registries, a.k.a. non-cancer controls
  - Derived from a 5% random sample of Medicare eligible persons



Will There Ever Be Patients With Cancer in the Control Population?

- There may be people in the non-cancer cases that have a cancer diagnosis
  - The person diagnosed with cancer prior to moving to the area
  - The registry may have missed the case
  - A misdiagnosis in claims

Tracking People Across Files

- For each cancer patient included in the SEER-Medicare data, there is a unique 10-digit ID known as the “regcase” that can be used to track a person across files.
- The first two digits are the registry number.
- Controls have similarly structured IDs

Files Included as Part of SEER-Medicare

Records are divided into several files:

Enrollment Files	Medicare Utilization Files	Other Useful Information
• PEDSF- SEER data and Medicare enrollment data • Sumdenom (non-cancer controls)	• MedPAR (hospital and skilled nursing care) • NCH/Carrier (physician and lab) • Hospital Outpatient (facility claims from clinics, same day surgery, etc.) • Hospice • Home Health • Durable Medical Equipment • Prescription drug information for beneficiaries with Part D	• Hospital file • CCW flag file • MDS (nursing home) • OASIS (home health care)

The SEER Data Included in SEER-Medicare

- SEER data are included in the Patient Entitlement and Diagnosis Summary File (PEDSF).
- Differs from SEER public access format-- has up to 10 incident cancer diagnoses per individual
- The PEDSF has flags with information about Medicare enrollment and HMO eligibility. This can be used to subset cases
- The PEDSF is also linked to SES data from the Census (zip code or census tract level)

Talking About the SEER-Medicare Data

Linkage Year	Newest Incident cases	Medicare claims data through	Released
2017	2014-2015	2016	2018
2015	2012-2013	2014	2016

- It is important to carefully check files to make sure that dates match your study question
  - Treatment coverage
  - Policy change
- There is always a lag in data availability:
  - Time for the SEER registries to identify and abstract data
  - Time for data to be submitted to NCI and linked with Medicare
    - Linkages done every 2 years
  - Creation of data files

Strengths of Using SEER-Medicare Data

- Include large numbers of cases- currently 1.8 million patients age 65+
- Include longitudinal measures of health care use from the time of Medicare coverage until death
- Span most clinical areas where health care is delivered
- Represent a diversity of geographic areas across the U.S.
- Are population-based and thus reflect “real world” practice
- Include data on multiple disease conditions- occurring before and after cancer diagnosis
- Include a cancer-free control group so can look at the baseline healthcare use of the elderly

Limitations of SEER-Medicare Data

Generalizability of the data:	Some services are not in the data:	Not a clinical database:	Data are Observational
• Under 65 population only includes the disabled/ESRD • Data are always a few years behind	• Non-covered services • Care not covered by the Medicare program - ex. VA or IHS • Care for Medicare patients in managed care except part D	• Reasons for tests are not known; • Results of tests not available • Do no know if treatment was offered, but not received • Symptoms and behaviors inconsistently reported	• Decisions about how to treat patients not random • May be hard to control for unobserved variation • You do not know what you do not know

Are the Data Hard to Use?

- Yes. The data are tricky…
- Most people have more than one of each type of Medicare record in a calendar year
- Many events result in multiple separate bills
- Most of the bills are divided into multiple sub-records for efficient data storage
- An understanding of Medicare payment policy maybe needed to make sense of what shows up
- You will need to create rules to translate bills into analyzable variables

What are Some Examples of Research Findings Based on SEER-Medicare Data?

- Higher rates of hip fracture after radiation therapy found for women with cervical and rectal cancer and for men with prostate cancer
- Higher rates of rectal cancer for men receiving RT for prostate cancer
- Reduction in use of androgen deprivation therapy for non-indicated use after payment reduction, but no change in use for indicated conditions

Questions?

Defining Your Study Population

Segment 2

SEER-Medicare Training 2019

Kristi Swanson, MS

How Do I Get Started?

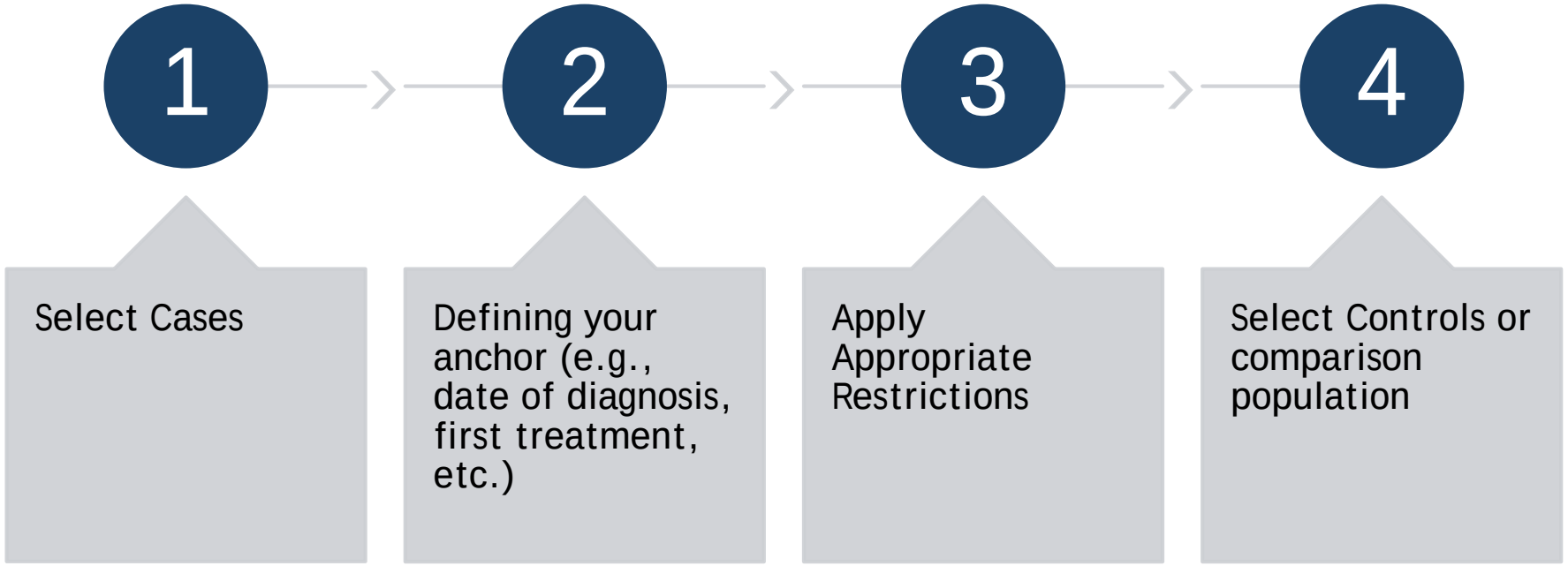
- Think about your research question and ask: “Who is the target population that fits that question?”
- Examples
  - Broad Research Question: “What are the effects of…..”
  - Population Definitions
    - CRC
    - 65 and older
    - Etc.
  - Questions to ask yourself
    - Any particular stages?
    - Etc.

SCHOOL OF PUBLIC HEALTH
UNIVERSITY OF MINNESOTA

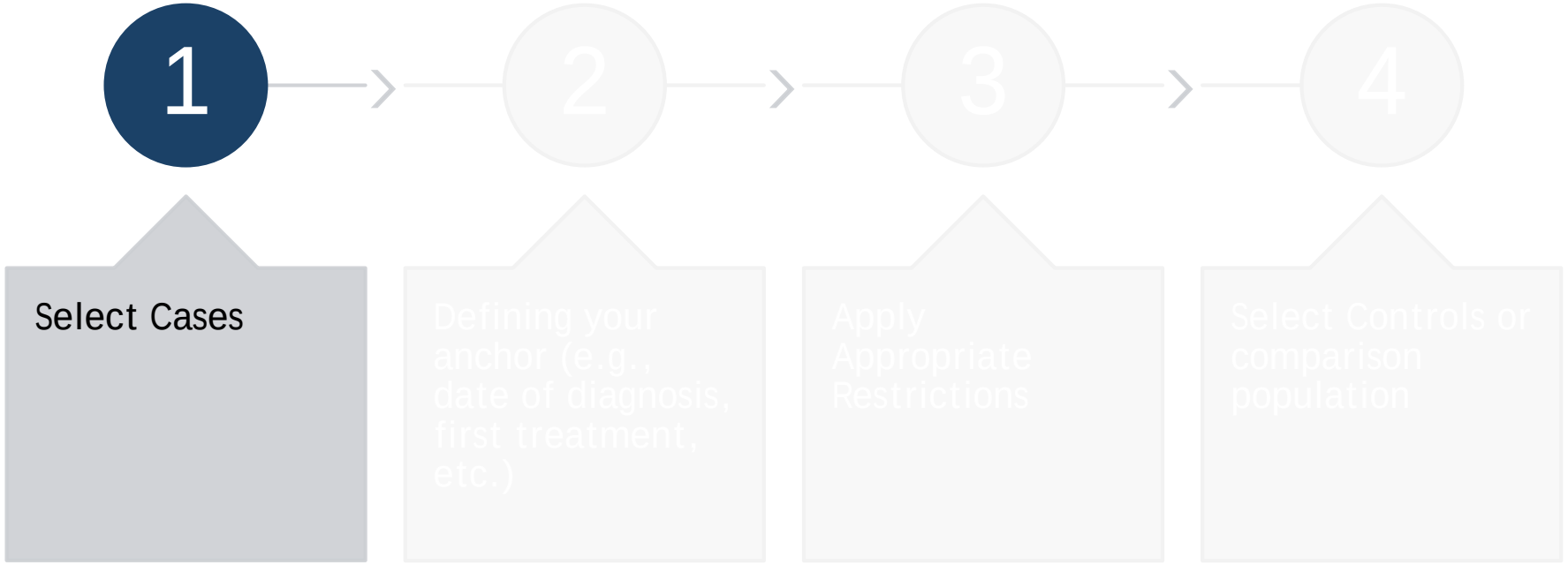
Important Things to Consider When Defining Your Study Population

- Getting the right sample for your study – Clinical and policy relevance
- Maximize your flexibility during analysis
- Generalizability
  - Keep publication in mind when developing your cohort and you will save a lot of time!
- Timeliness
- Practice Changes Over Time

Steps



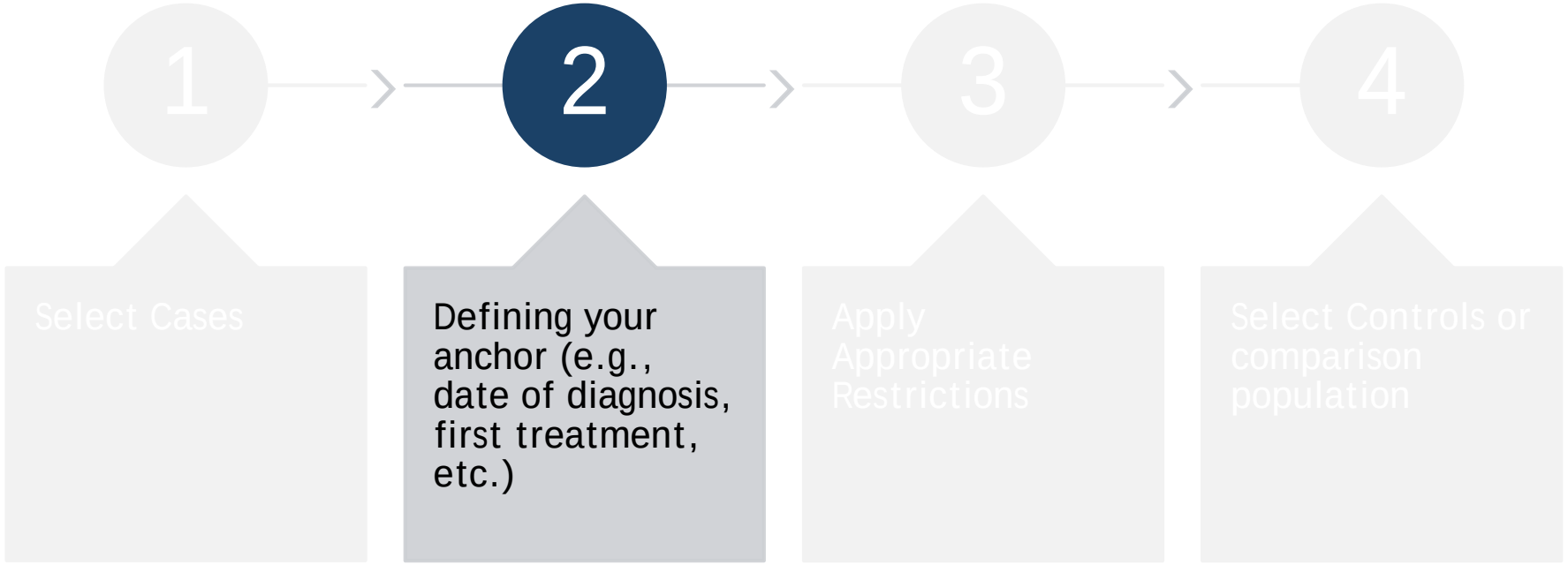
Step 1



Select Cases

- When you get the cases, you will need to utilize various fields to identify your target population
- This is typically done using the PEDSF File
- Can identify patients using:
  - Cancer Site (Documentation)
  - Histology (Documentation)
  - Stage
  - Age at diagnosis
  - Date of diagnosis – this can be tricky
    - PEDSF only gives the month and year of diagnosis

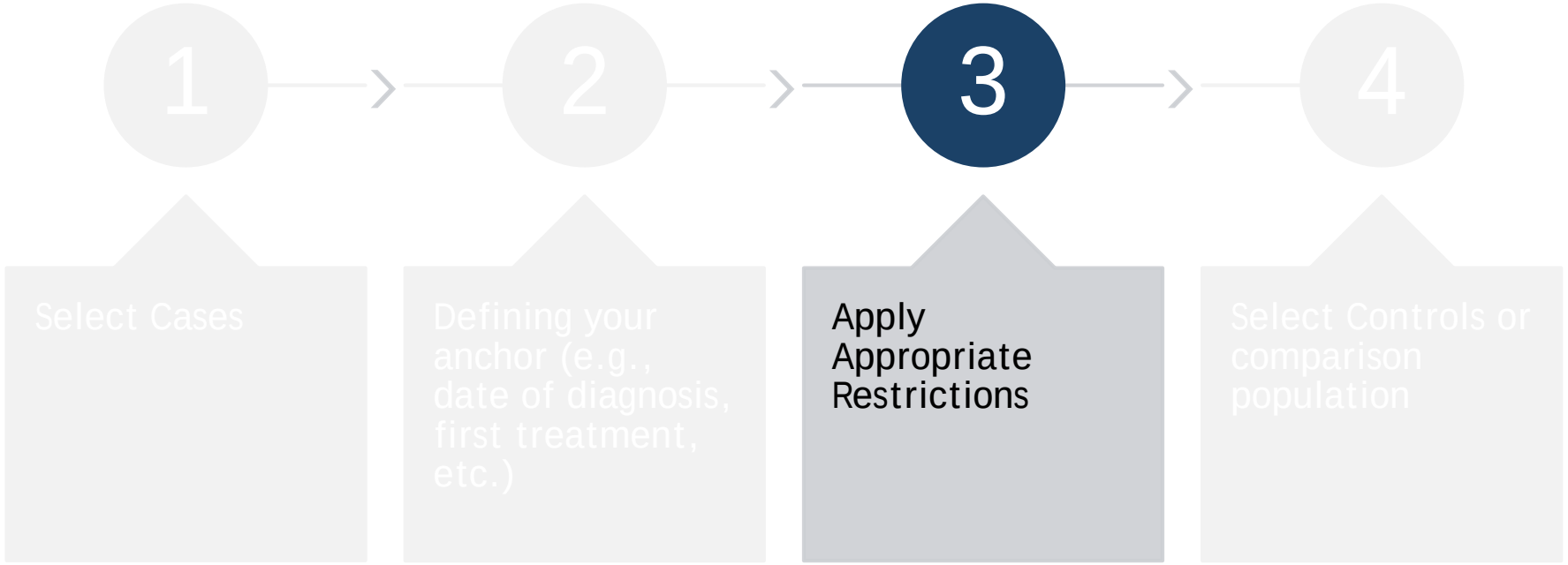
Step 2



Defining Your Anchor!

- Where is your starting point for your cases?
  - Necessary for survival/time-to-event analysis
  - Also needed to set end-point for pre-diagnosis history
- Examples
  - Date of Diagnosis
    - SEER gives month and year of diagnosis
    - From Medicare you can find dates like:
      - Date of first treatment
      - Date of screening test
      - First date the cancer diagnosis showed up
- Setting anchor dates can be tricky, especially if you want to examine a comparator population that doesn't have the disease or doesn't receive the treatment of interest
  - What do you anchor on in this group??

Step 3



Apply SEER-Based Restrictions

- Once you have this population identified, you may want to refine your cohort using applicable exclusion criteria
- Histology (if you didn't already use to identify your cases)
- Missing Workup (unknown stage, grade, histology…)
- Diagnosed by Death certificate or autopsy (i.e., never treated)
- Sequence number
  - First and only, first of many, second, etc.
- Minimum survival (did they live long enough to be treated?)

Histology, Stage, etc.

- Remember that SEER collects detailed information about the cancer
- These fields may help you hone in on your cohort of interest
- Some useful cancer related fields may include:
  - Histology
  - Stage
  - Some tumor markers
  - Lymph node testing and positivity
  - Grade
  - Size
  - Extent of disease

Exclusions Based on Reporting Source (i.e., How the Registry Learned About the Cancer)

- Reporting source options – these are rank ordered:
  - '1' – Hospital
  - '3' – Laboratory
  - '4' – Private Doctor
  - '5' – Nursing/Convalescent
  - '6' – Autopsy
  - '7' – Death Certificate
- Those cases identified by autopsy or death certificate are often excluded when evaluating treatment. Depending on the study, you may also want to exclude nursing home as reporting source.

Exclusions Based on Sequence Number

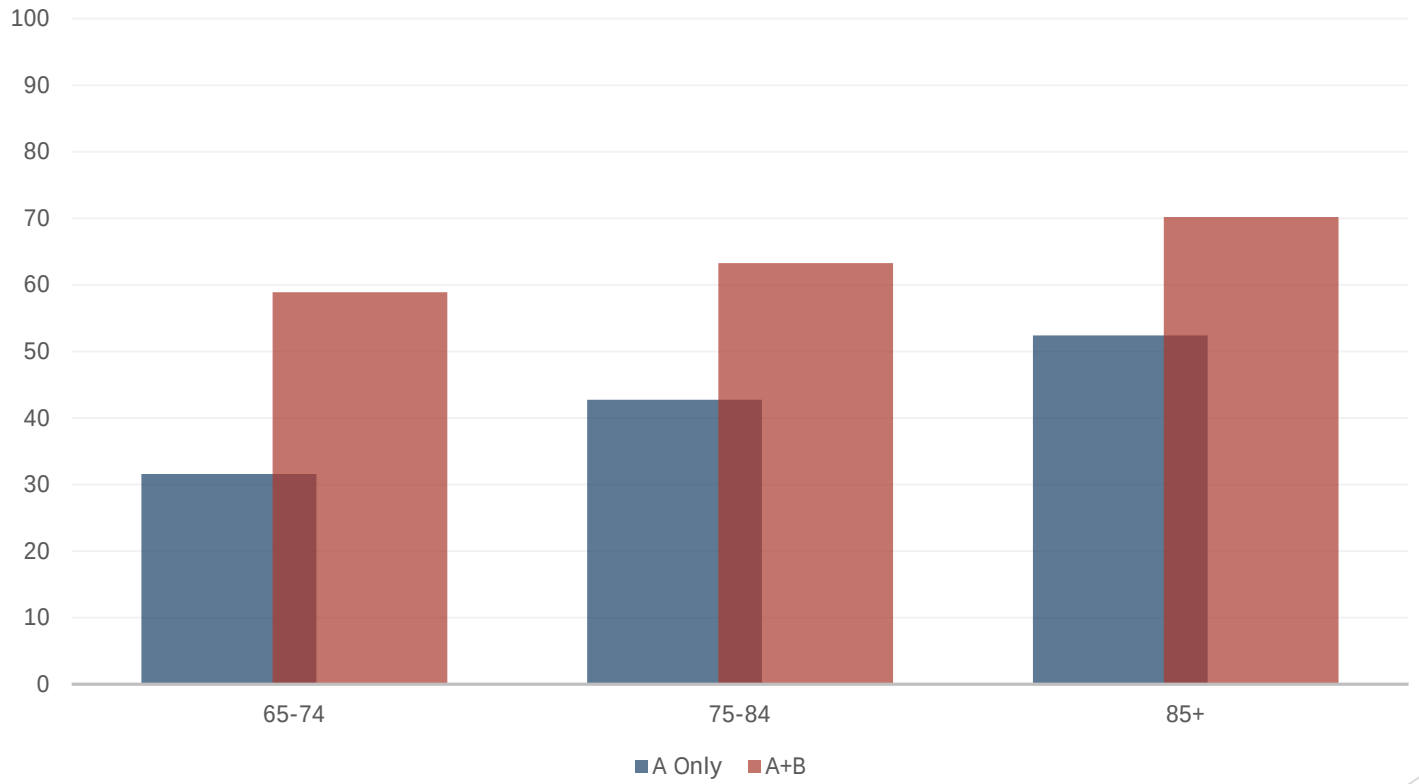
- The sequence number describes the number and order of all reportable malignant, in situ, benign, and borderline primary tumors, which occur over the lifetime of a patient.
- For example, an individual with a single reportable malignant neoplasm will have a sequence code for the first neoplasm of ’00’ = First and Only Cancer
- If that same individual, instead, subsequently had other cancers diagnosed, the sequence code for the first neoplasm would be ‘01’ = First of Many Cancers
- And so on….
  - E.g., ‘02’ = Second of Many Cancers
- Researchers typically limit analyses based on this field (e.g., to First and Only or First of Many Cancers)

Apply Restrictions Based on Medicare Coverage

- If you need to measure receipt of care or intend to follow patients over time (e.g., to assess treatment, survival, etc.)
  - You must have Part A for: Hospital, SNF, Hospice
  - You must have Part B for: ER, Outpatient, DME, IV chemotherapies, radiation therapy, etc.
  - Part D: Chemotherapy for which there is no IV equivalent or metabolite, other prescription medications for cancer or non-cancer indications
- Researchers often limit analyses to those with both Parts A and B

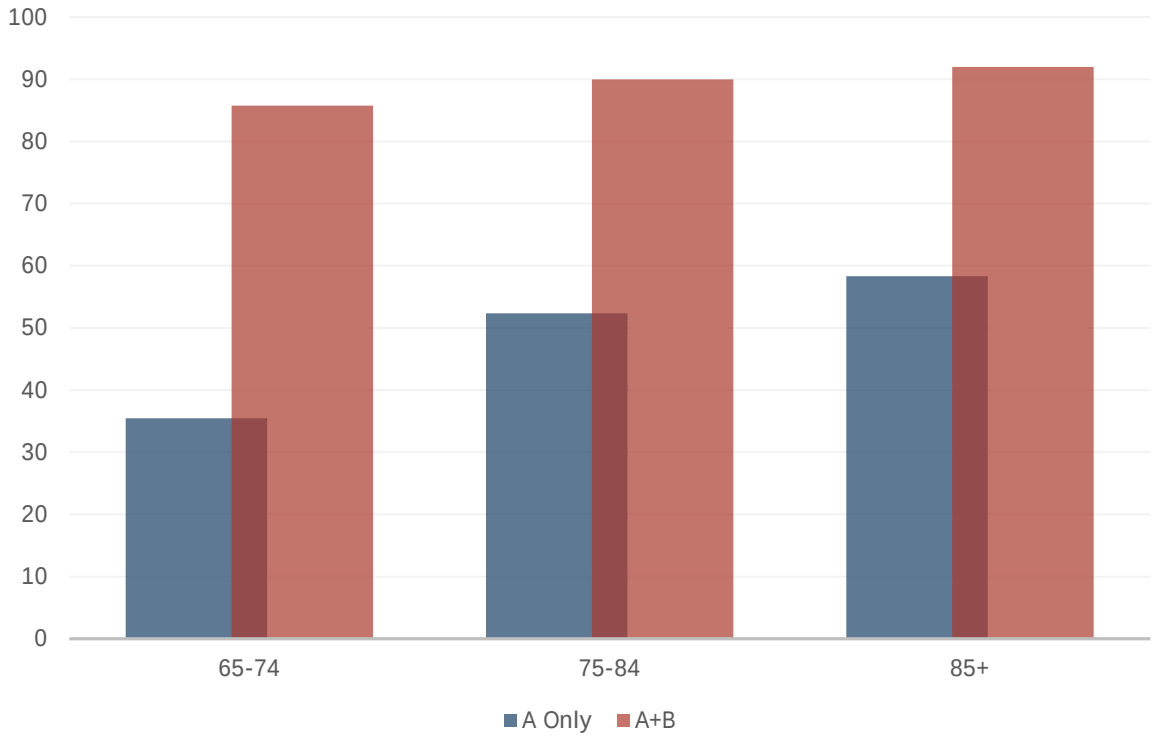
Why Might We Need Both Parts A and B?: Percent Of Cases Hospitalized by Age and Coverage

- Persons with A-only coverage probably have incomplete claims, even for Part A services:



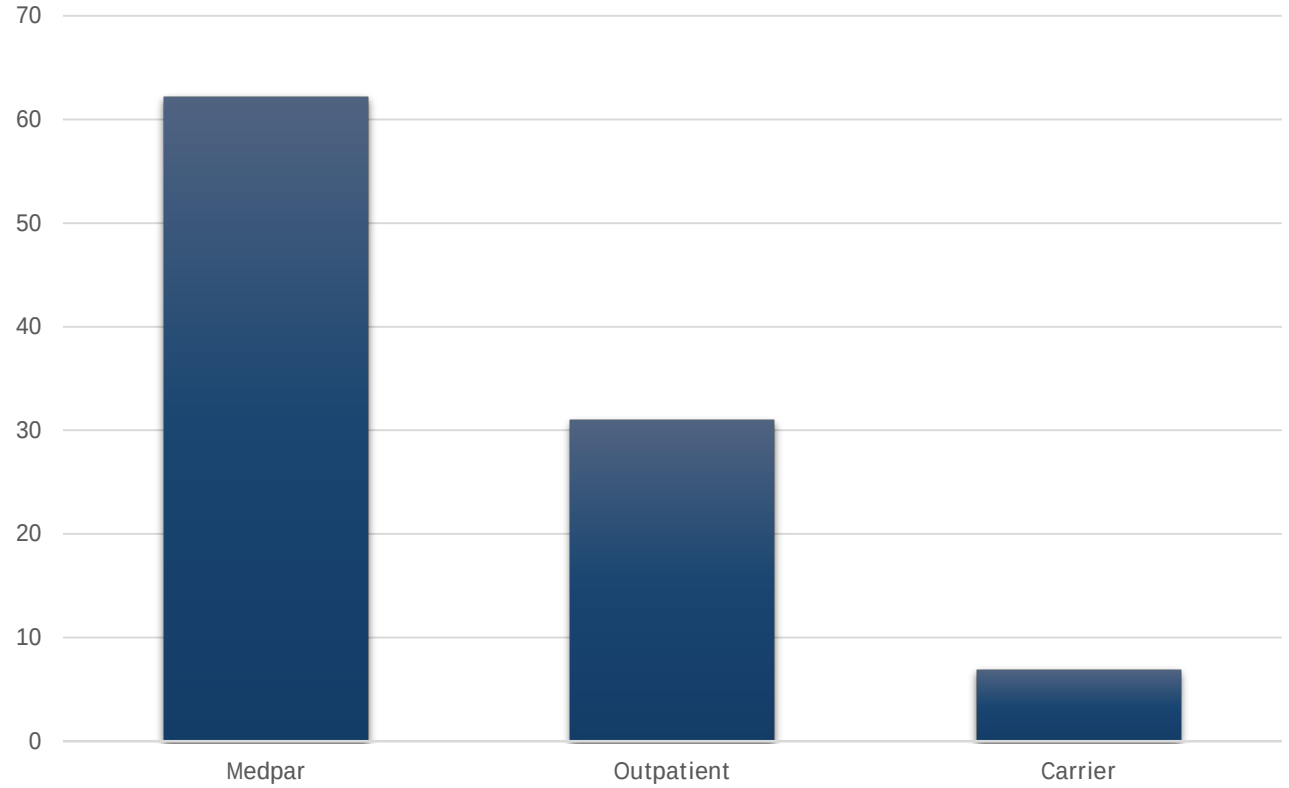
This Pattern is Seen With Patients ‘Known’ to Have Treatment

- Hospitalization rates for 2004-2013 incident colorectal cancer cases treated with surgery – SEER-Medicare



Coverage Example: Services Don't Always Happen Where You Think They Do

- Consider a breast cancer cohort
- Mastectomies happen in the hospitals…..or do they?



Coverage Example Part 2: What Do We Need to Look at Comprehensive Breast Cancer Care?

- To look comprehensively at breast cancer treatment

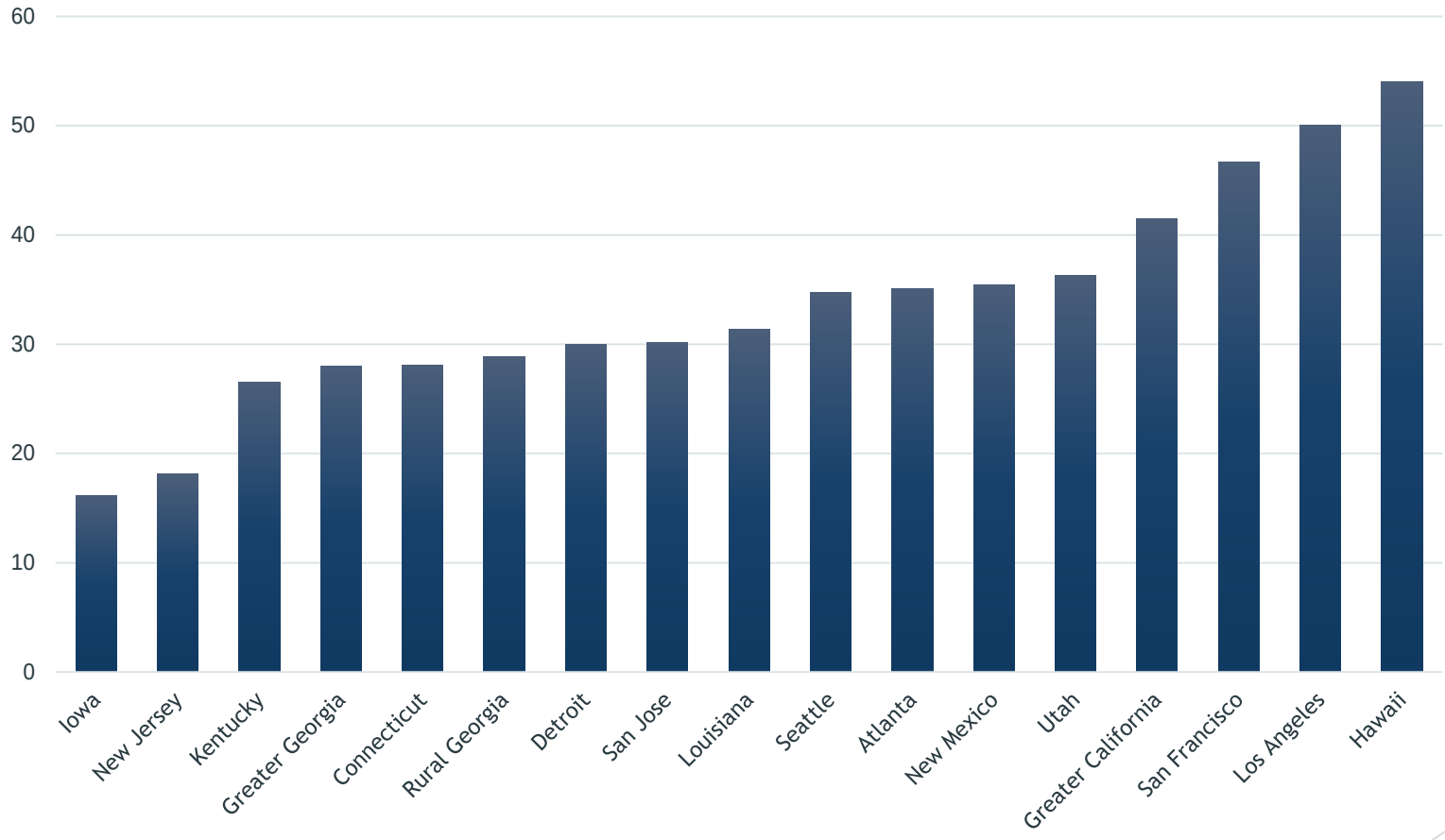
Service	Part A	Part B	Part D
Mastectomy (hospital, MD)	X	X	
BCS	X	X	
RT		X	
Herceptin (IV chemotherapy)		X	
Tamoxifen or Aromatase Inhibitors			X

- Take away: It is confusing to mix an A-only cohort with an A+B (depending on location of service). So A+B often makes more sense and results in a more coherent manuscript.

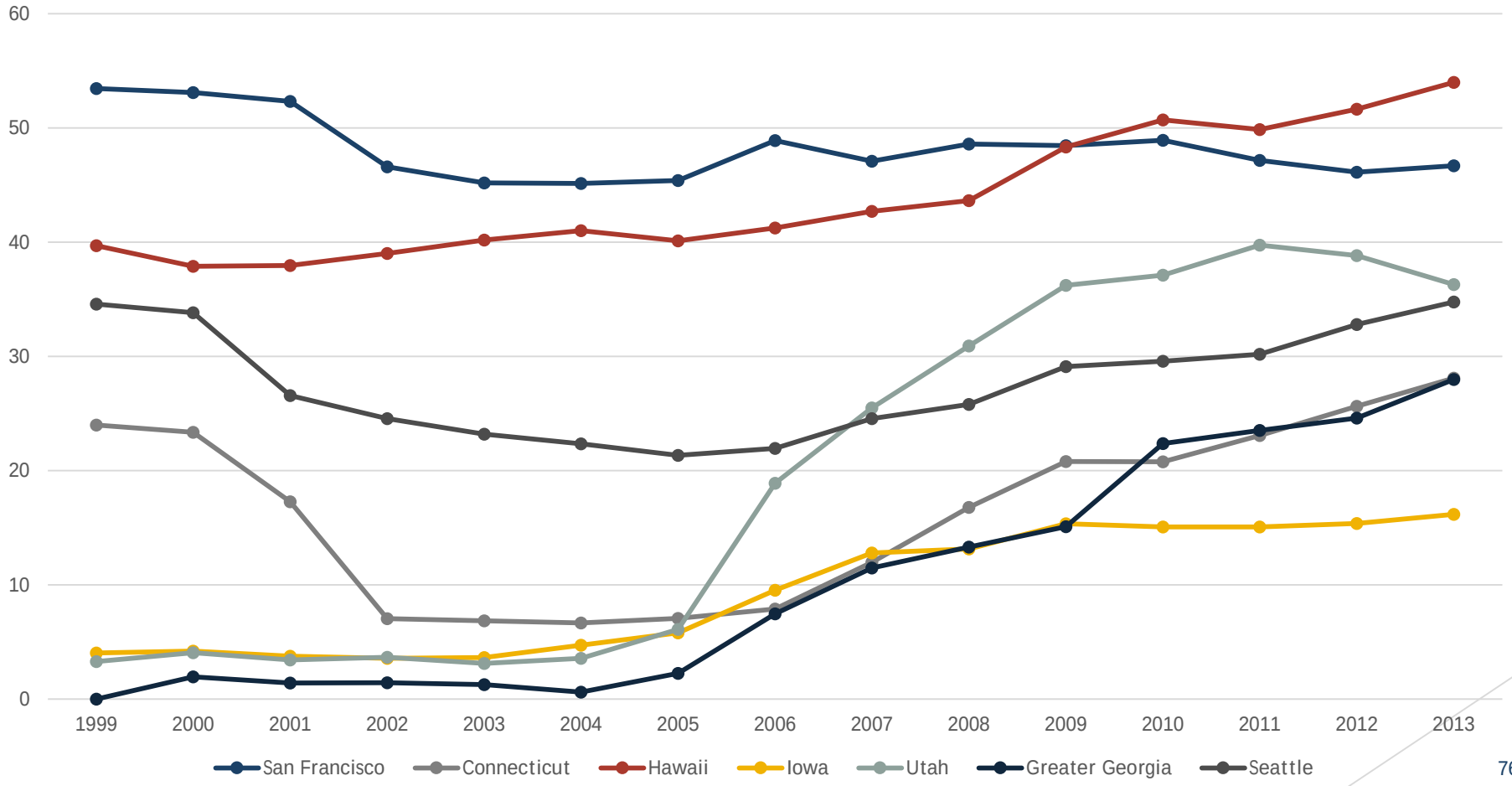
Additional Medicare Based Restrictions: Managed Care

- Enrollees in Medicare managed care plans receive all Medicare covered benefits
- However, the managed care plans do not submit or transmit detailed claims to the Medicare program other than hospice for A or B benefits
- Thus, with the exception of treatment information provided by SEER, there is no information about Part A or B health care services other than hospice for managed care enrollees
- Part D information is available for MC enrollees

Managed Care Enrollment Varies Across Registries: Cancers Diagnosed In 2014 and Enrolled in Medicare at the Time Of Diagnosis



Medicare Managed Care Enrollment Varies Over Time and by Registry: 1999-2013

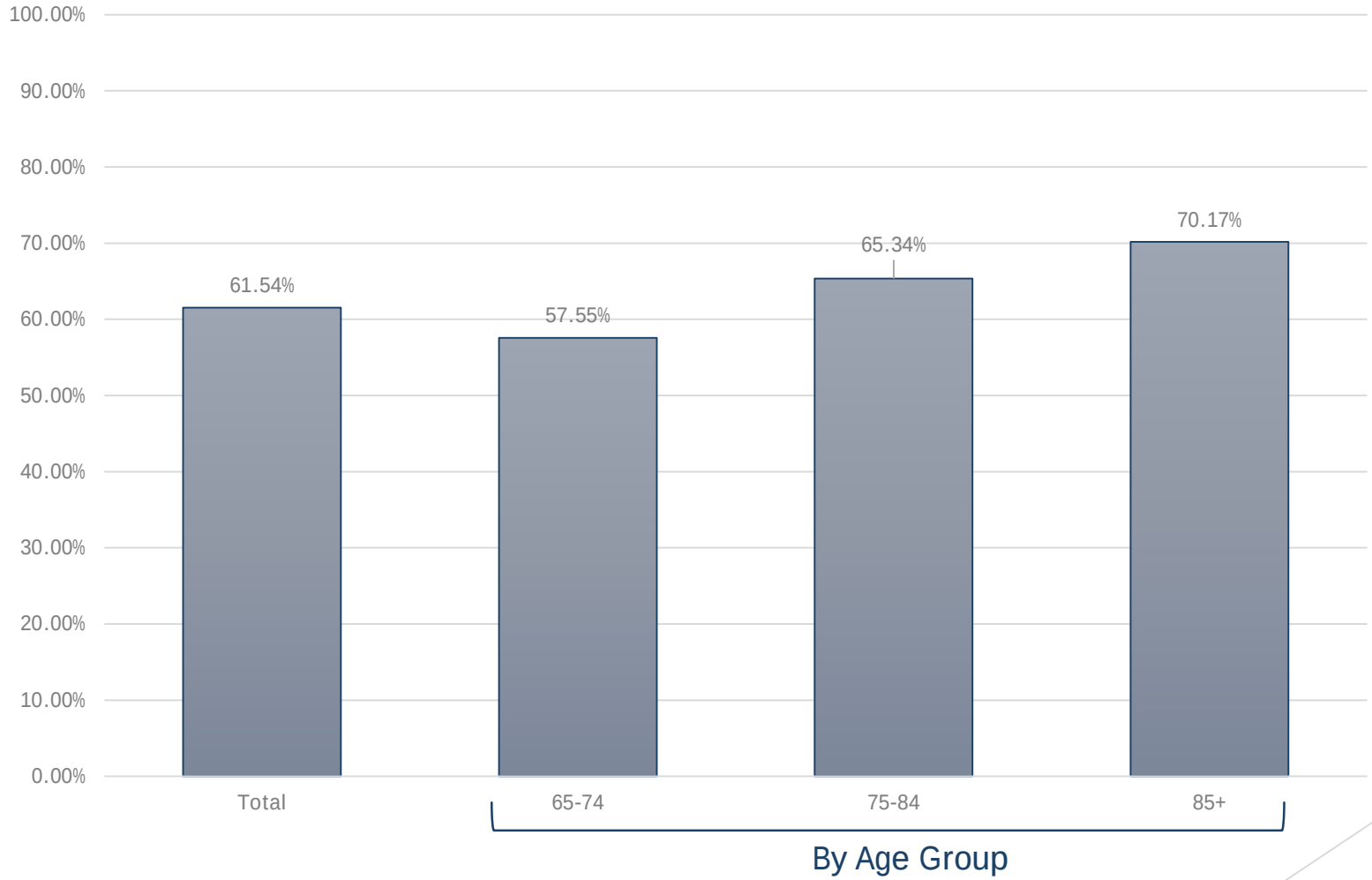


SCHOOL OF PUBLIC HEALTH
UNIVERSITY OF MINNESOTA

Those “Likely to Have Claims”

- Many Medicare-based analyses are limited to ‘elderly persons likely to have claims at the time of diagnosis’
- The idea is that we limit to people where we can be most confident that we will see all their care in the Medicare data
- To identify this group include:
  - Persons ages 65 or older and enrolled in Medicare at the time of diagnosis
  - Persons not enrolled in Medicare managed care (remember, for managed care we only have submits hospice claims)
  - Persons who have equal and continuous Parts A and B coverage during the entire period of observations (or until censoring event) (more on this next)

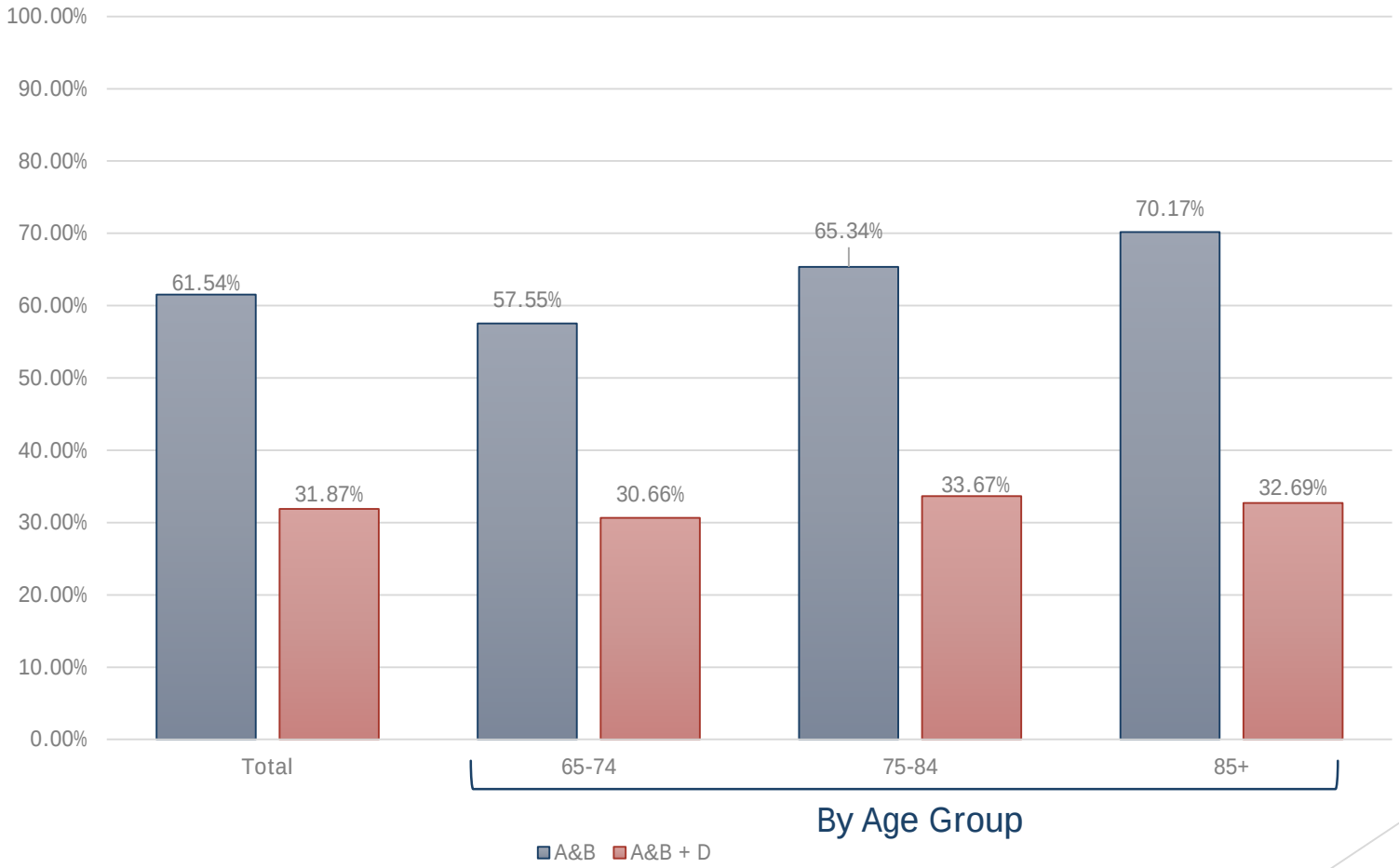
Percent of Breast Cancer Cases Likely to Have Complete Claims (Cases Diagnosed from 2007 to 2013 With Coverage for a Year After Diagnosis or Until Death)



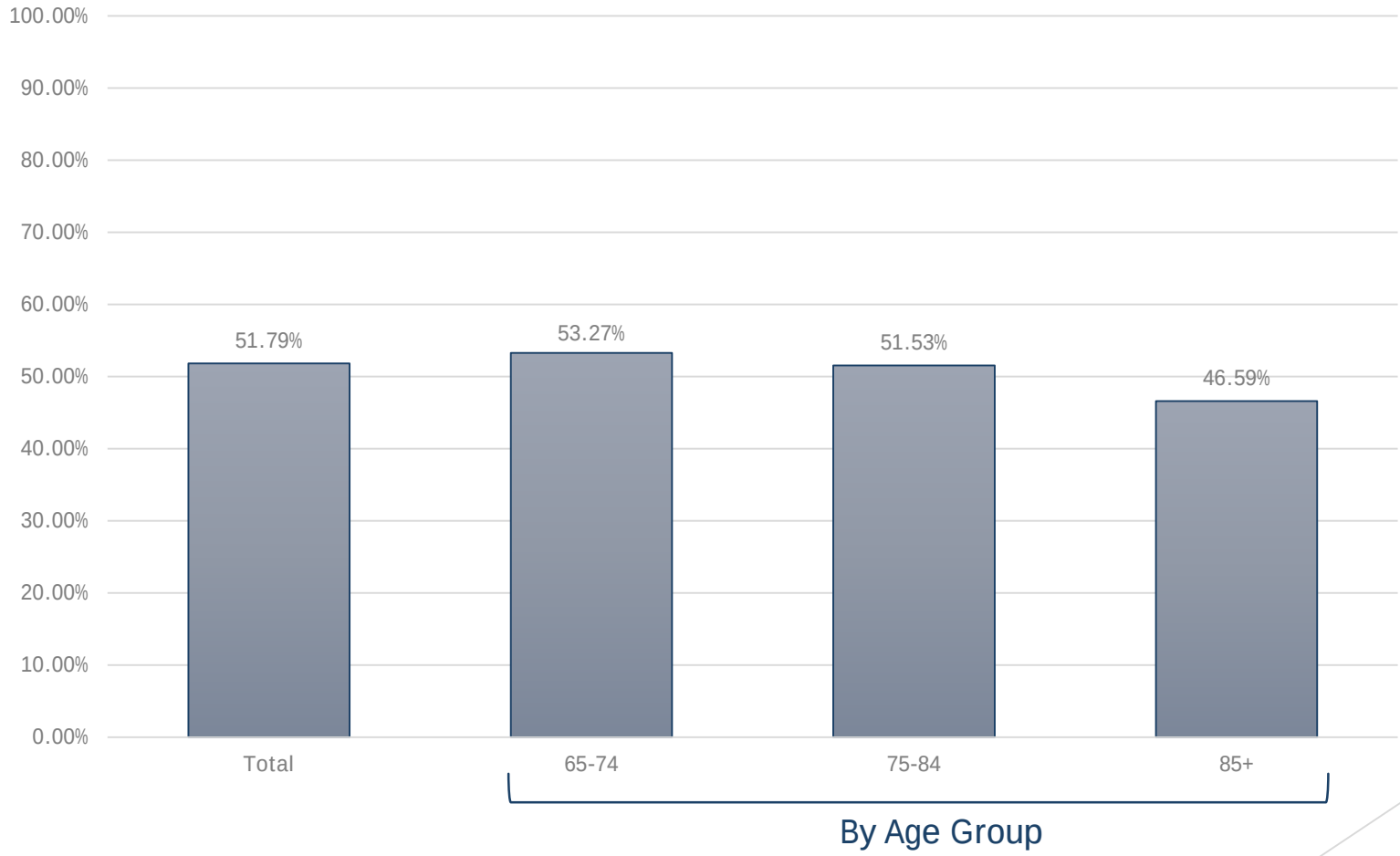
Don’t Forget About Part D

- Remember that Part D is prescription drug coverage
- If your research question aims to examine prescription drug therapies, then you need to also require Part D coverage
- Something to be aware of:
  - Certain chemotherapy drugs (and other groups of drugs) may be covered under Part B or part D depending on how they are available and administered, as well as based on the indication for which they are prescribed
  - We will talk about this more in Segment 4!

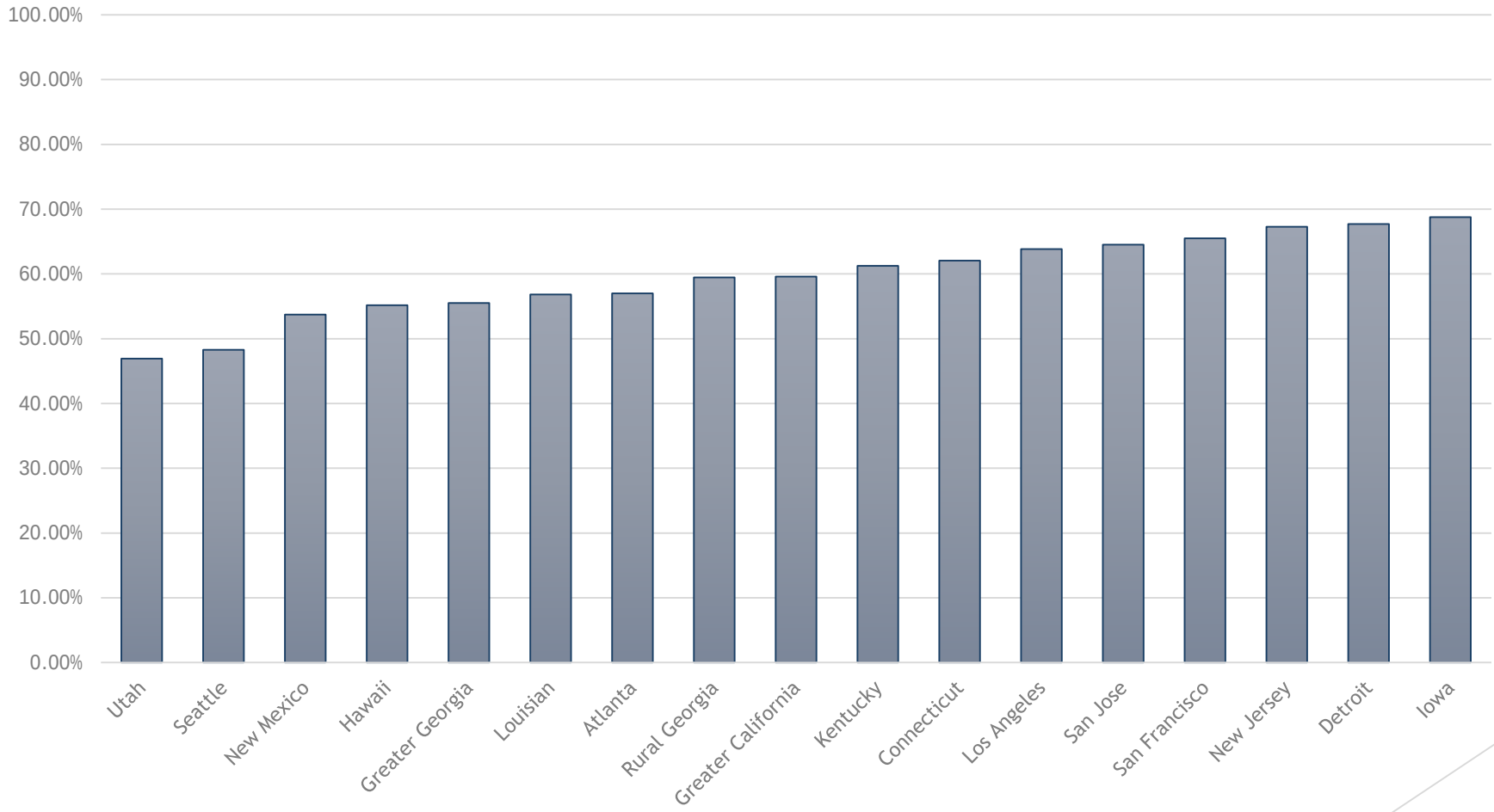
Adding Part D Coverage



Percent of A&B (Likely to Have Complete Claims) With Part D Coverage



Part D Enrollment Among Likely to Have Complete Claims: By Registry in 2014



Summary of Common Limits for Studies of Treatment

- From SEER:
  - Histology, stage, tumor characteristics
  - Not diagnosed on death certificate or autopsy
  - First cancer (sequence number)
  - Survive at least X time
- From Medicare:
  - Medicare A+B entire time, no HMO enrollment (likely to have complete claims)
  - Sometimes limit upper age of cases
  - Sometimes limit lower age of cases so pre-diagnosis information is available (e.g., baseline comorbidity)
  - Part D if needed

How Long Do I Require Coverage For?

- Often, researchers will require that these conditions be met for at least 6 months (or some other time period, such as at least one year for comorbidity identification) prior to diagnosis
- The length for which you require coverage after diagnosis depends on your research question
  - How long will you be looking for the care plus a little more…
- However, it is important to understand how length of coverage restrictions impact your study

Length of Enrollment – It's a Balancing Act

- Think about where you would expect to see the care taking place and how long you will need to look for it
- So, let's say I want to look at cases diagnosed between 2004 and 2015 with coverage at the time of diagnosis

Length of Coverage	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
At Diagnosis	Sample Population											
At diagnosis and for 1 year prior												
At diagnosis and for 1-year prior and 1-year post												

Length of Enrollment – It's a Balancing Act

- Now, what if we say, we also want them to have coverage for a year prior? How about for a year post, as well?

Length of Coverage	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
At Diagnosis												
At diagnosis and for 1 year prior	Sample Population											
At diagnosis and for 1-year prior and 1-year post												

Length of Coverage	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
At Diagnosis												
At diagnosis and for 1 year prior												
At diagnosis and for 1-year prior and 1-year post	Sample Population											

Be Aware!!

- By requiring this additional coverage, you have effectively changed your sample population.
- You say “I want cases diagnosed from 2004 through 2015 who are 65 and older with Medicare coverage for one year prior to diagnosis and for a full year after diagnosis or until death”
- What you will end up with is a cohort of those 66 years and older (by requiring coverage for a year prior)
- Coverage indicators only go so far in the data. For the 2016 linkage, they go through December of 2015. Therefore, anyone that didn’t die during the study period, must have been diagnosed no later than December of 2014 to meet your coverage requirements.
  - Will you get anyone diagnosed in 2015?
  - Depends on if you require a full year of coverage (i.e., didn’t die in the year after) or a full year of coverage OR until death (in which case your cases diagnosed in 2015 also died in 2015)

Length of Enrollment – It's a Balancing Act

- What if I add a Part D requirement, as well?

Length of Coverage	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
At Diagnosis												
At diagnosis and for 1 year prior												
At diagnosis and for 1-year prior and 1-year post				Sample Population								

- What happened?
  - Remember that Part D data doesn't start being available until 2007. PLUS, the coverage indicators in the 2016 linkage only go through 2014.

Length of Enrollment – It's a Balancing Act

- How about all of the above **PLUS** 5 years of coverage?

Length of Coverage	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
At Diagnosis												
At diagnosis and for 1 year prior												
At diagnosis and for 1-year prior and 1-year post				Sample Population								

- Moral of the story….think thoughtfully about how your study design impacts your sample size due to length of enrollment requirements

Let's Look at a Concrete Example: Impact of Restrictions on Cohort Size?

- Diagnosed with breast cancer from 2004 to 2013: 363,209(100%)
- First and only cancer and not diagnosed by autopsy or death certificate: 293,926 (80.9%)
- Stage restrictions – Stages I-IV: 230,608(63.5%)
- Tumor grade is known (e.g., well differentiated, moderately differentiated, etc.): 215,157(59.2%)
- Age 65 or older: 133,168(36.7%)
- A&B at diagnosis and for year after diagnosis: 124,008(34.1%)
- No Managed care: 84,502(23.3%)
- Part D at diagnosis and for year after diagnosis: 31,827(8.8%)
  - Final cohort = 31,827 (8.8%)

- **BUT……**

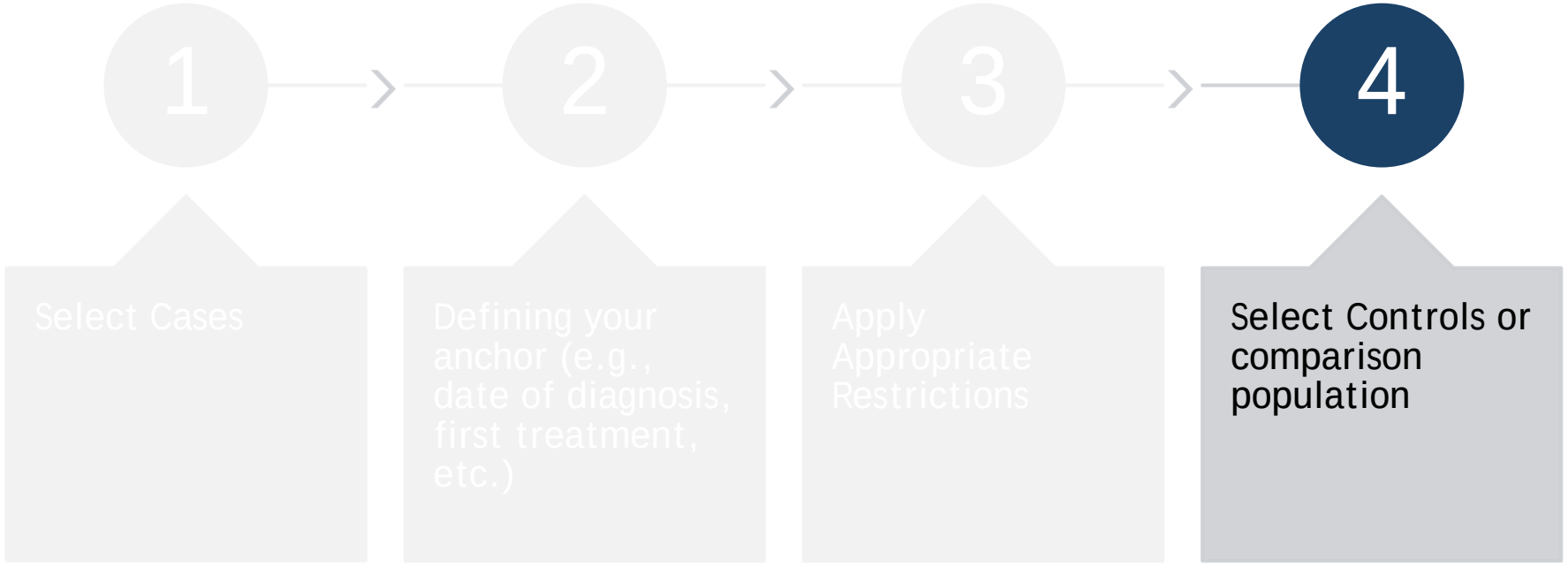
Think About Your Starting Point

- Keep in mind what requiring Part D enrollment implies!! (effectively gives you a cohort diagnosed from 2007 to 2014)
- So let's run the numbers again
  - Diagnosed with breast cancer from **2007** to 2013: 248,524(100%)
  - First and only cancer and not diagnosed by autopsy or death certificate: 205,543 (82.7%)
  - Stage restrictions – Stages I-IV: 161,969(65.2%)
  - Tumor grade is known (e.g., well differentiated, moderately differentiated, etc.): 151,815(61.1%)
  - Age 65 or older: 98,651(39.7%)
  - A&B at diagnosis and for year after diagnosis: 91,585(36.9%)
  - No Managed care: 60,301(24.3%)
  - Part D at diagnosis and for year after diagnosis: 31,827(12.8%)
    - Final cohort = 31,827 (12.8%)
- Be careful when selecting projects…

Key Takeaways

- Numbers can get small very quickly
- As more inclusion-exclusion criteria are used, the potential sample size may shrink
- Solutions
  - Relax constraints (looser criteria)
  - Consider treating missing as a category
  - Add more years of cases
  - Stick with the available number of well-defined cases

Step 4



Should You Use a Comparison Population? Benefits

- Placing the results in context:
  - Do people with RT live longer than people without?
  - Are hospitalization rates different by…
  - Does race explain treatment?
  - What are the non-clinical/tumor factors associated with treatment choice?

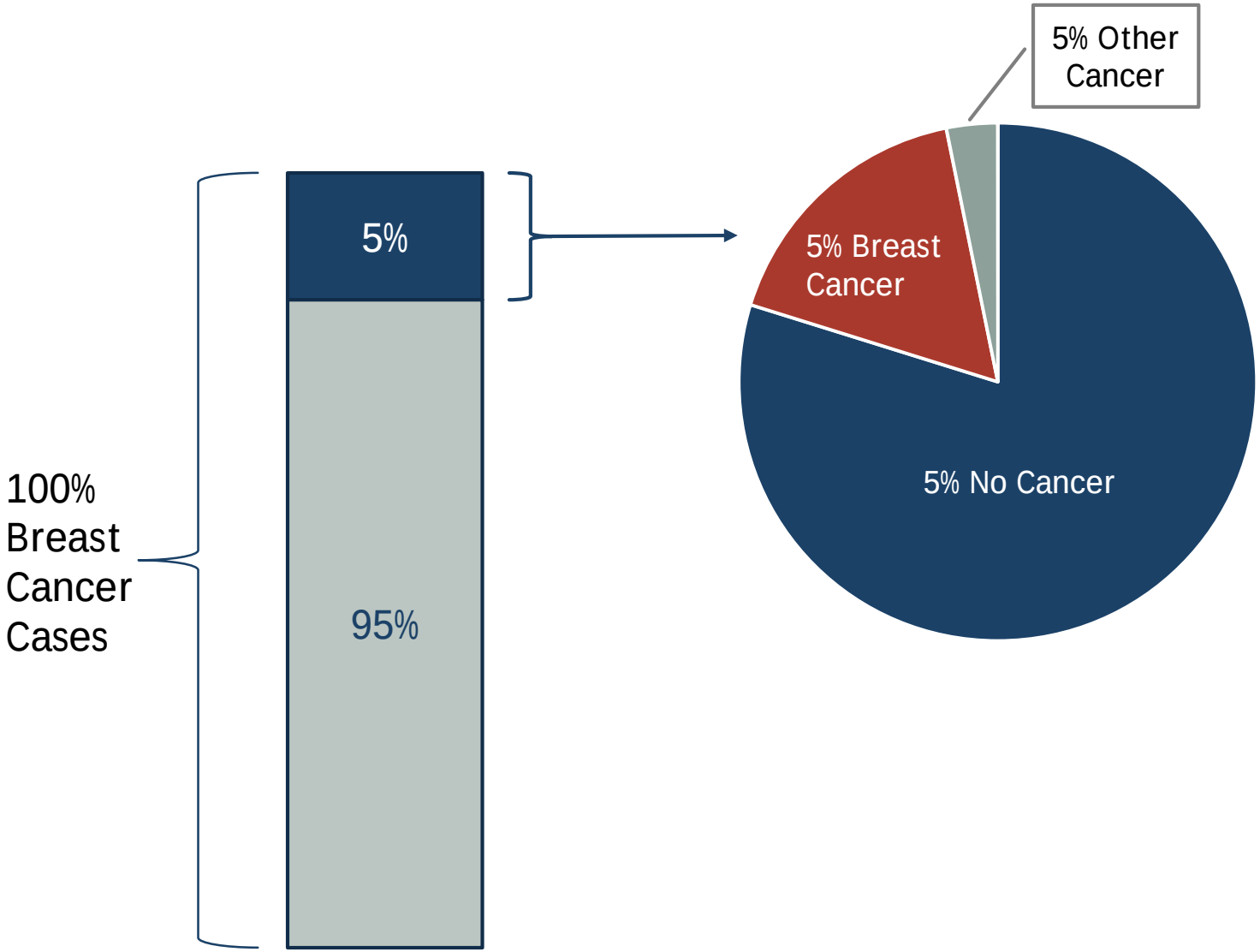
What Might Be Considered When Picking a Comparison Population

Biology – is it possible that the disease-outcome link reflects common biology between disease and outcome?	Pick a control with the same disease
Health care system effects	Pick controls with the same access to health care
Treatment effects	Pick controls who otherwise would have equal likelihood of getting the outcome but for the treatment

Options for Source of Comparison Population

- Same cancer, different demographic, stage or treatment
  - Select individuals who are equally eligible to have outcome
- Non-cancer and 'other cancer' control populations
  - 5% sample of Medicare beneficiaries residing in SEER areas
  - Non-cancer sample has never had a SEER-reported cancer
  - 'Other cancer' sample is part of Medicare 5% sample and has had a SEER-reported cancer, but it is different cancer than those in the case cohort

Two Parts Create a Whole



Should You Match on Year of Diagnosis?

- Decision may vary depending on the study, research questions, etc. but secular trends may introduce confounding because…
  - SEER staging has changed over time
  - Clinically, technology changes over time for treatment, pathology, etc.
  - New guidelines
  - ICD coding changed in 2015 from ICD-9 to ICD-10
  - CPT codes change periodically
  - Medicare benefits, payments and covered services change
  - Drugs may go off patent…

When Selecting the Control Population it is Important to Consider

- Are they truly comparable but for the factor of my study?
  - There may be some people with cancer in the control population
    - Linkage failure
    - Prevalent cancer case moved to SEER area or before registry started
    - Mis-diagnosis
    - SEER missed the case
  - Controls are selected because Medicare considers them to live in a SEER area. Cases are selected because SEER considers them to live in a SEER area
  - Studies of Race (and other demographic factors) that include the control population should use Medicare not SEER variables for both cases and controls

When Selecting Control Populations (and, Perhaps Case Populations) Consider…

- Are both cases and controls eligible for my outcome?
  - Cause of death information is not available for non-cancer control populations
- How much time in Medicare is needed prior to diagnosis?
  - 1 year is typical, but that also may limit the study population – no one diagnosed prior to age 66?
- What is the mechanism for an effect? How much time will you need for follow-up?
  - if it will take 3 years for the effect to be seen, then the last case cannot be enrolled less than 3 years prior to the last Medicare data

Summary

- Using SEER and Medicare data together is a powerful way to select a study sample that has comparable demographic and cancer characteristics
- It is important to balance comparability and sample size needs
- Be thoughtful in your exclusion criteria and understand how it impacts your sample population
- Make sure you implement coverage requirements appropriately (e.g., have part B when you need it, have part D when you need it, etc.)
  - Limit to those "likely to have claims"
- Remember that coverage requirements (and possibly other restrictions) vary over time and across registries which may have implications for your study

Questions?

Break

We will return at 11:15 ET / 10:15 CT

Baseline Measures

Segment 3

SEER-Medicare Training 2019
Stephanie Jarosek, PhD

SCHOOL OF PUBLIC HEALTH
UNIVERSITY OF MINNESOTA

Understanding Your Cohort

- Cohort vs comparison group(s)
  - Describe how they compare at baseline
  - Decide how to handle differences (matching, propensity scores, stratification, etc.)
- May be an iterative process
  - You may learn things about your cohort or comparison group that require you to revisit your cohort inclusion/exclusion criteria

Example: Prostate Cancer

- Urinary adverse events
  - Men who received treatment for prostate cancer
  - Non-cancer controls

Table 1 – Unweighted demographic characteristics of non-cancer control group and prostate cancer cohort stratified by treatment group

	Treatment group						
	Control	EBRT	BT	BT+EBRT	RP	RP+EBRT	Cryotherapy
Number	144 816	44 318	14 259	11 835	26 790	1557	2,115
Age at treatment (%)							
65-69	36.74	17.55	25.91	25.29	51.00	51.12	22.30
70-74	30.83	37.21	40.26	40.56	40.37	39.82	34.06
75-79	19.18	32.64	26.54	27.02	7.97	[b]	29.90
80-84	13.25	12.59	7.29	7.13	0.66	[b]	13.75
Race (%)							
White	83.33	83.10	88.70	84.36	86.70	84.59	82.95
Black	6.62	9.99	6.67	9.51	6.55	6.23	10.54
Hispanic	2.50	1.78	1.02	1.61	1.95	2.44	2.35
Asian	3.95	2.99	2.03	2.57	2.28	4.11	2.26
Other unknown	3.60	2.13	1.59	1.94	2.52	2.63	1.90
Median income (%)							
Q1	27.08	22.53	20.54	17.92	19.38	18.43	28.00
Q2	25.82	24.22	21.97	20.23	25.30	26.72	24.69
Q3	24.16	25.44	26.99	28.07	26.75	26.91	25.10
Q4	22.94	27.81	30.50	33.78	28.56	27.94	22.21
High school completion (%)							
Q1	27.36	23.50	21.01	20.81	18.30	20.36	27.23
Q2	25.54	25.05	23.70	22.61	22.55	21.77	24.88
Q3	24.58	25.83	26.32	26.13	26.47	27.17	25.64
Q4	22.52	25.63	28.97	30.44	32.69	30.70	22.25
Charlson score (%)							
0	69.11	62.11	65.57	62.92	68.28	65.19	54.50
1	19.68	24.36	23.68	25.56	23.45	25.05	29.13
2	6.85	8.43	7.21	7.72	5.98	7.19	10.36
3+	4.35	5.09	3.55	3.80	2.29	2.57	6.02
Clinical T stage (%)							
1	N/A	37.85	50.89	38.85	37.11	31.28	42.20
2x[a]	N/A	16.83	17.15	16.89	15.01	13.55	14.88
2y	N/A	11.32	5.91	13.38	10.40	12.14	9.14
2z	N/A	25.85	23.78	25.40	28.31	23.96	28.22
3 or 4	N/A	5.10	0.36	3.77	7.27	17.53	2.89
Unknown	N/A	3.05	1.91	1.71	1.90	1.54	2.67
WHO grade (%)							
1	N/A	4.92	3.45	2.18	4.45	2.12	1.09
2	N/A	57.81	80.01	52.49	63.85	40.98	52.92

Most Baseline Measures are Found in the PEDSF (Patient Entitlement and Diagnosis Summary File)

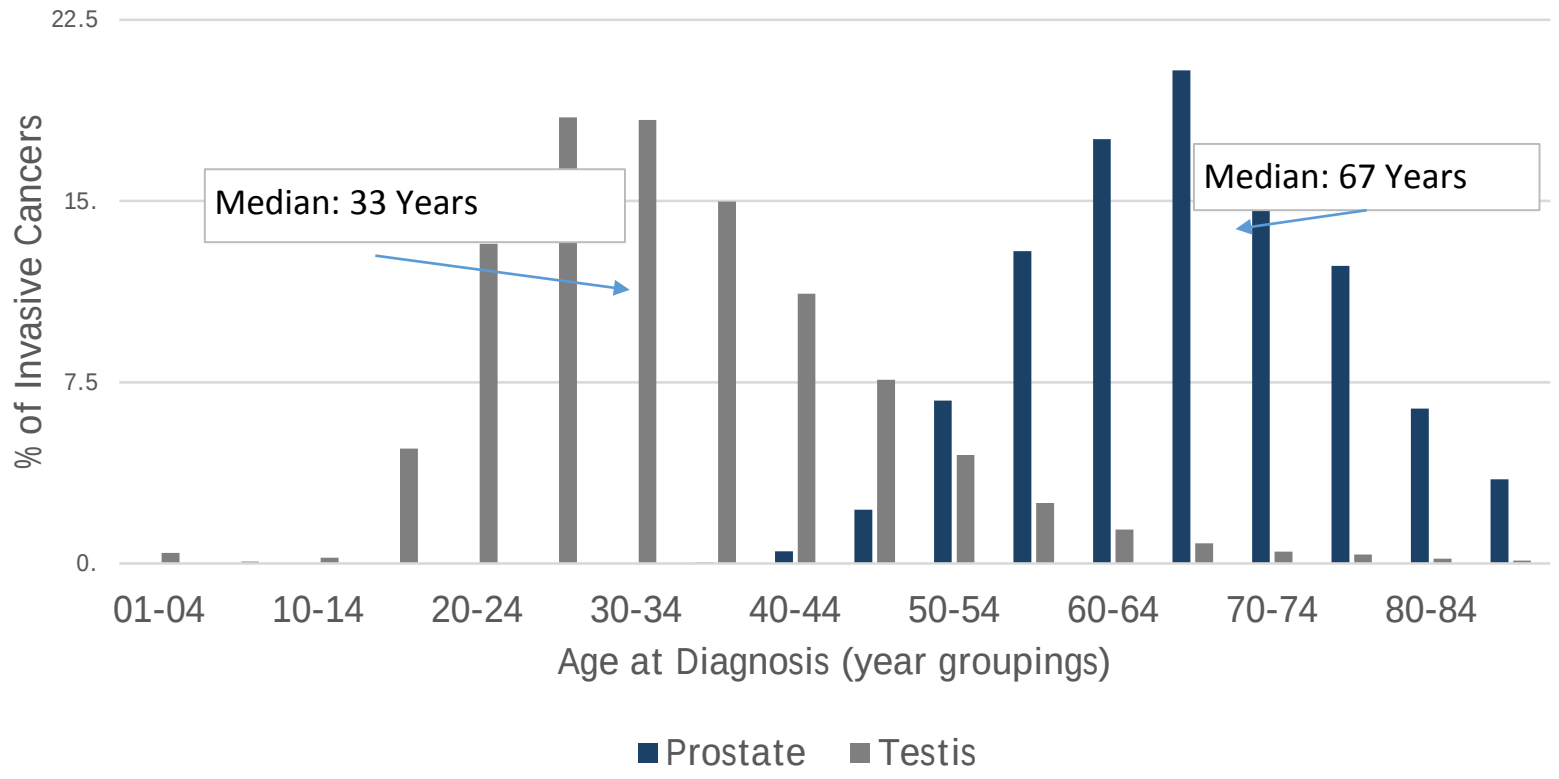
- Age
- Race
- Sex
- Registry
- Other geography
  - HSA, urban/rural
- Cancer diagnosis information
  - Extent of disease
  - Cancer specific information

Others May Need to Be Linked and/or Summarized

- Comorbidity measures
  - Need to be summarized from claims and linked via patient identifier
- Census tract or zip code level variables (often used for socioeconomic status)
  - Linked to PEDSF
- Information for non-cancer controls
  - Medicare enrollment and demographic information available in SUMDENOM (summarized denominator) file

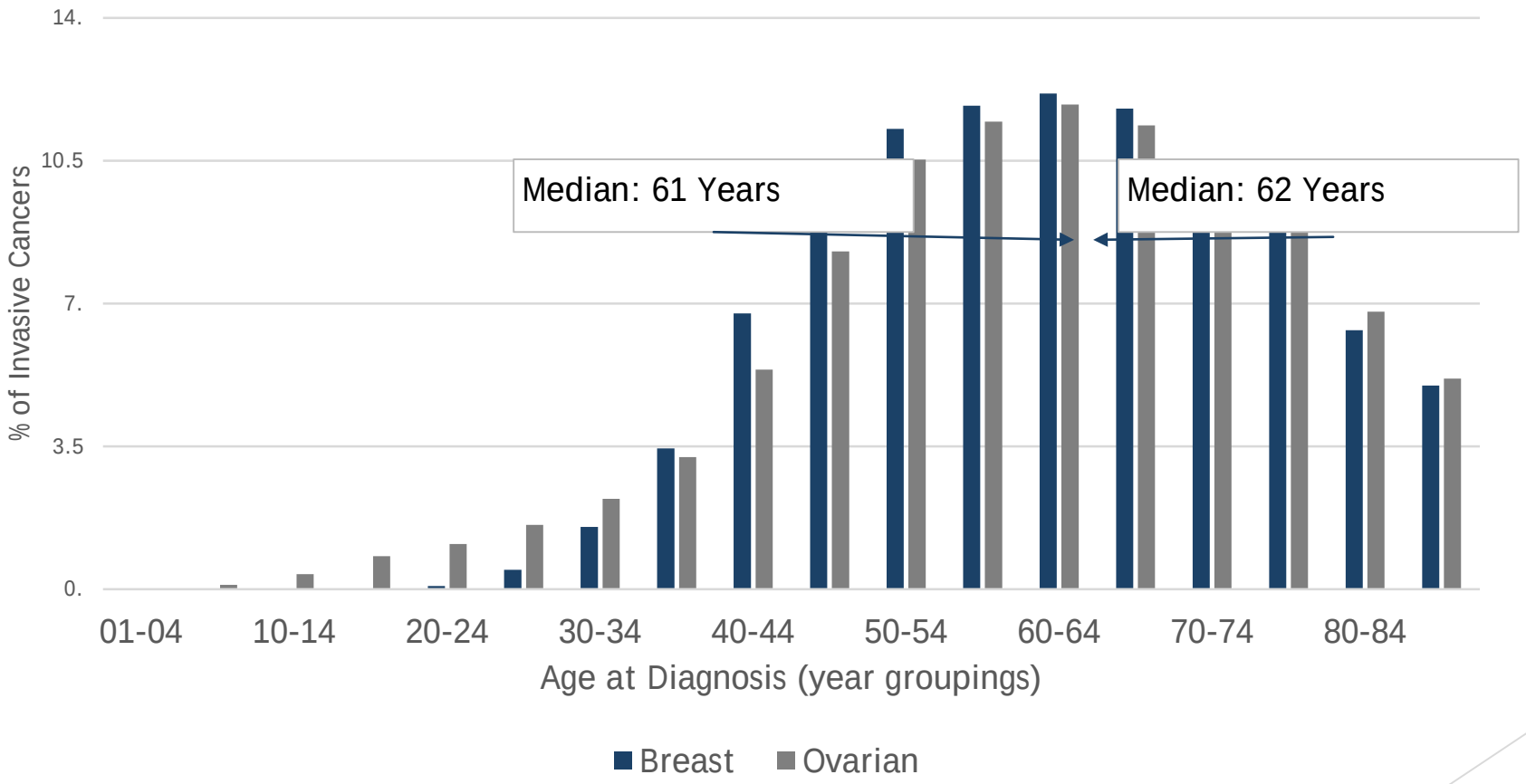
Age at Diagnosis

Prostate vs Testicular Cancer



Age at Diagnosis

Female Breast vs Ovarian Cancer



Age at Diagnosis

- Increased age may reduce the likelihood of surgical procedures and/or aggressiveness of treatment

Table 1 – Unweighted demographic characteristics of non-cancer control group and prostate cancer cohort stratified by treatment group

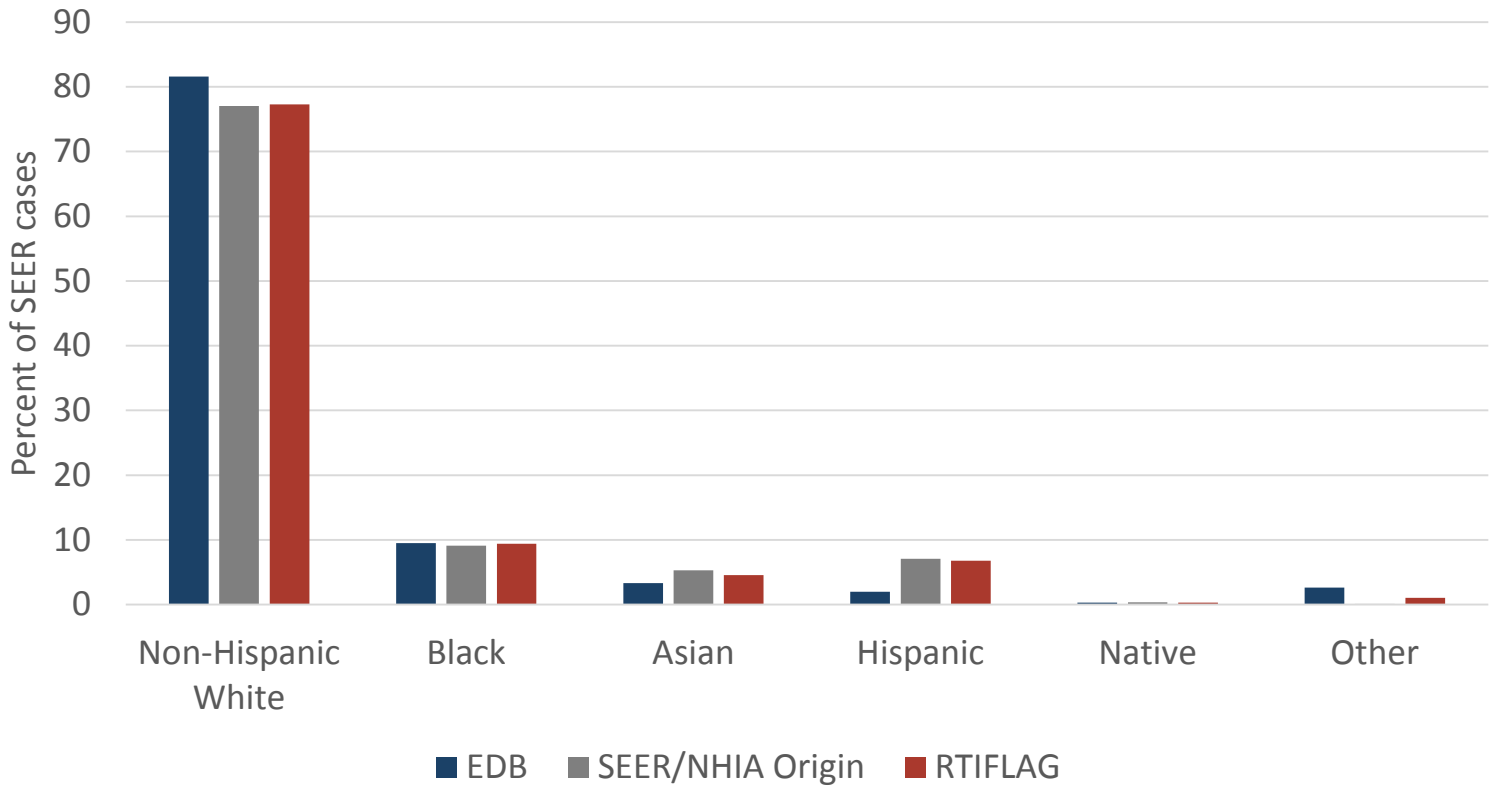
	Treatment group						
	Control	EBRT	BT	BT+EBRT	RP	RP+EBRT	Cryotherapy
Number	144 816	44 318	14 259	11 835	26 790	1557	2,115
Age at treatment (%)							
65-69	36.74	17.55	25.91	25.29	51.00	51.12	22.30
70-74	30.83	37.21	40.26	40.56	40.37	39.82	34.06
75-79	19.18	32.64	26.54	27.02	7.97	[b]	29.90
80-84	13.25	12.59	7.29	7.13	0.66	[b]	13.75
Race (%)							
White	83.33	83.10	88.70	84.36	86.70	84.59	82.95
Black	6.62	9.99	6.67	9.51	6.55	6.23	10.54
Hispanic	2.50	1.78	1.02	1.61	1.95	2.44	2.35
Asian	3.95	2.99	2.03	2.57	2.28	4.11	2.26
Other unknown	3.60	2.13	1.59	1.94	2.52	2.63	1.90
Median income (%)							
Q1	27.08	22.53	20.54	17.92	19.38	18.43	28.00
Q2	25.82	24.22	21.97	20.23	25.30	26.72	24.69
Q3	24.16	25.44	26.99	28.07	26.75	26.91	25.10
Q4	22.94	27.81	30.50	33.78	28.56	27.94	22.21
High school completion (%)							
Q1	27.36	23.50	21.01	20.81	18.30	20.36	27.23
Q2	25.54	25.05	23.70	22.61	22.55	21.77	24.88
Q3	24.58	25.83	26.32	26.13	26.47	27.17	25.64
Q4	22.52	25.63	28.97	30.44	32.69	30.70	22.25
Charlson score (%)							
0	69.11	62.11	65.57	62.92	68.28	65.19	54.50
1	19.68	24.36	23.68	25.56	23.45	25.05	29.13
2	6.85	8.43	7.21	7.72	5.98	7.19	10.36
3+	4.35	5.09	3.55	3.80	2.29	2.57	6.02
Clinical T stage (%)							
1	N/A	37.85	50.89	38.85	37.11	31.28	42.20
2x[a]	N/A	16.83	17.15	16.89	15.01	13.55	14.88
2y	N/A	11.32	5.91	13.38	10.40	12.14	9.14
2z	N/A	25.85	23.78	25.40	28.31	23.96	28.22
3 or 4	N/A	5.10	0.36	3.77	7.27	17.53	2.89
Unknown	N/A	3.05	1.91	1.71	1.90	1.54	2.67
WHO grade (%)							
1	N/A	4.92	3.45	2.18	4.45	2.12	1.09
2	N/A	57.81	80.01	52.49	63.85	40.98	52.92

Race/Ethnicity

Several race variables are available in the PEDSF

- Medicare (*=added in 1994)
  - White
  - Black
  - Asian/PI*
  - Hispanic*
  - North American native/AK Native*
  - Unknown
  - Other
- RTI race
  - Hispanic surname algorithm
- SEER
  - Race recode Y
    - White, Black, American Indian/Alaska Native, Asian or Pacific Islander
  - Race recode A
  - Race/Ethnicity
  - Origin
  - NHIA Hispanic Origin
  - Derived API race
    - Asian/Pacific Islander surname algorithm

Medicare vs SEER Race/Ethnicity 2007-2014



SCHOOL OF PUBLIC HEALTH
UNIVERSITY OF MINNESOTA

Race/Ethnicity

- Sometimes there are systematic differences in treatment groups by race
  - Black men are more represented in external beam radiation

Table 1 – Unweighted demographic characteristics of non-cancer control group and prostate cancer cohort stratified by treatment group

	Treatment group						
	Control	EBRT	BT	BT+EBRT	RP	RP+EBRT	Cryotherapy
Number	144 816	44 318	14 259	11 835	26 790	1557	2,115
Age at treatment (%)							
65-69	36.74	17.55	25.91	25.29	51.00	51.12	22.30
70-74	30.83	37.21	40.26	40.56	40.37	39.82	34.06
75-79	19.18	32.64	26.54	27.02	7.97	[b]	29.90
80-84	13.25	12.59	7.29	7.13	0.66	[b]	13.75
Race (%)							
White	83.33	83.10	88.70	84.36	86.70	84.59	82.95
Black	6.62	9.99	6.67	9.51	6.55	6.23	10.54
Hispanic	2.50	1.78	1.02	1.61	1.95	2.44	2.35
Asian	3.95	2.99	2.03	2.57	2.28	4.11	2.26
Other unknown	3.60	2.13	1.59	1.94	2.52	2.63	1.90
Median income (%)							
Q1	27.08	22.53	20.54	17.92	19.38	18.43	28.00
Q2	25.82	24.22	21.97	20.23	25.30	26.72	24.69
Q3	24.16	25.44	26.99	28.07	26.75	26.91	25.10
Q4	22.94	27.81	30.50	33.78	28.56	27.94	22.21
High school completion (%)							
Q1	27.36	23.50	21.01	20.81	18.30	20.36	27.23
Q2	25.54	25.05	23.70	22.61	22.55	21.77	24.88
Q3	24.58	25.83	26.32	26.13	26.47	27.17	25.64
Q4	22.52	25.63	28.97	30.44	32.69	30.70	22.25
Charlson score (%)							
0	69.11	62.11	65.57	62.92	68.28	65.19	54.50
1	19.68	24.36	23.68	25.56	23.45	25.05	29.13
2	6.85	8.43	7.21	7.72	5.98	7.19	10.36
3+	4.35	5.09	3.55	3.80	2.29	2.57	6.02
Clinical T stage (%)							
1	N/A	37.85	50.89	38.85	37.11	31.28	42.20
2x[a]	N/A	16.83	17.15	16.89	15.01	13.55	14.88
2y	N/A	11.32	5.91	13.38	10.40	12.14	9.14
2z	N/A	25.85	23.78	25.40	28.31	23.96	28.22
3 or 4	N/A	5.10	0.36	3.77	7.27	17.53	2.89
Unknown	N/A	3.05	1.91	1.71	1.90	1.54	2.67
WHO grade (%)							
1	N/A	4.92	3.45	2.18	4.45	2.12	1.09
2	N/A	57.81	80.01	52.49	63.85	40.98	52.92

Socioeconomic Status

- Zip code or census tract level variables
- Data derived from the decennial Census (1990, 2000) or the American Community Survey
- Variables are generally continuous
  - ex: Median income,% with high school education
- Some are age and race specific
- Missingness
  - ~5% missing Zip Code SES information
  - Excluded here

Table 1 – Unweighted demographic characteristics of non-cancer control group and prostate cancer cohort stratified by treatment group

	Treatment group						
	Control	EBRT	BT	BT+EBRT	RP	RP+EBRT	Cryotherapy
Number	144 816	44 318	14 259	11 835	26 790	1557	2,115
Age at treatment (%)							
65-69	36.74	17.55	25.91	25.29	51.00	51.12	22.30
70-74	30.83	37.21	40.26	40.56	40.37	39.82	34.06
75-79	19.18	32.64	26.54	27.02	7.97	[b]	29.90
80-84	13.25	12.59	7.29	7.13	0.66	[b]	13.75
Race (%)							
White	83.33	83.10	88.70	84.36	86.70	84.59	82.95
Black	6.62	9.99	6.67	9.51	6.55	6.23	10.54
Hispanic	2.50	1.78	1.02	1.61	1.95	2.44	2.35
Asian	3.95	2.99	2.03	2.57	2.28	4.11	2.26
Other unknown	3.60	2.13	1.59	1.94	2.52	2.63	1.90
Median income (%)							
Q1	27.08	22.53	20.54	17.92	19.38	18.43	28.00
Q2	25.82	24.22	21.97	20.23	25.30	26.72	24.69
Q3	24.16	25.44	26.99	28.07	26.75	26.91	25.10
Q4	22.94	27.81	30.50	33.78	28.56	27.94	22.21
High school completion (%)							
Q1	27.36	23.50	21.01	20.81	18.30	20.36	27.23
Q2	25.54	25.05	23.70	22.61	22.55	21.77	24.88
Q3	24.58	25.83	26.32	26.13	26.47	27.17	25.64
Q4	22.52	25.63	28.97	30.44	32.69	30.70	22.25
Charlson score (%)							
0	69.11	62.11	65.57	62.92	68.28	65.19	54.50
1	19.68	24.36	23.68	25.56	23.45	25.05	29.13
2	6.85	8.43	7.21	7.72	5.98	7.19	10.36
3+	4.35	5.09	3.55	3.80	2.29	2.57	6.02
Clinical T stage (%)							
1	N/A	37.85	50.89	38.85	37.11	31.28	42.20
2x[a]	N/A	16.83	17.15	16.89	15.01	13.55	14.88
2y	N/A	11.32	5.91	13.38	10.40	12.14	9.14
2z	N/A	25.85	23.78	25.40	28.31	23.96	28.22
3 or 4	N/A	5.10	0.36	3.77	7.27	17.53	2.89
Unknown	N/A	3.05	1.91	1.71	1.90	1.54	2.67
WHO grade (%)							
1	N/A	4.92	3.45	2.18	4.45	2.12	1.09
2	N/A	57.81	80.01	52.49	63.85	40.98	52.92

Other SES Information

- Medicaid enrollment
  - Partial duals—receive Medicare premium and copayment support
  - Full duals—receive Medicaid benefits
  - Variables
    - State buy in
      - Available in all years;
      - Formatted differently in PEDSF than in data from CMS
        - Codes 1, 2 and 3 indicate the receipt of some support
      - Not a measure of poverty
        - many people in poverty do not receive assistance
    - Dual enrollment
      - Available in 2007+ in Part D Denominator

Geography

- Registry
- Health Service Area
  - National Center for Health Statistics (NCHS)—a single county or cluster of contiguous counties which are relatively self-contained with respect to hospital care.
  - NCI modified—all counties in one HSA are from one state and/or SEER registry
    - i.e., no HSAs cross state or SEER registries boundaries
- Rurality
- Zip code variables
  - Restricted
  - Used to measure distance traveled to treatment

Rurality

- RUCA--Rural-urban commuting area (2000, 2010)
  - 4 levels
- Urban-rural indicator (2000, 2010)
  - 4 levels
- Rural-urban continuum (1993, 2003, 2013)
  - 9 levels

Comorbidity

- Definitions (van den Akker et al., JCE, 1998)
  - Comorbidity: additional diseases beyond the condition under study
  - Multimorbidity: any occurrence of multiple coexisting conditions or diseases
- Predicts mortality
- Impacts:
  - Treatment choices
  - Post-surgical complications
  - Costs

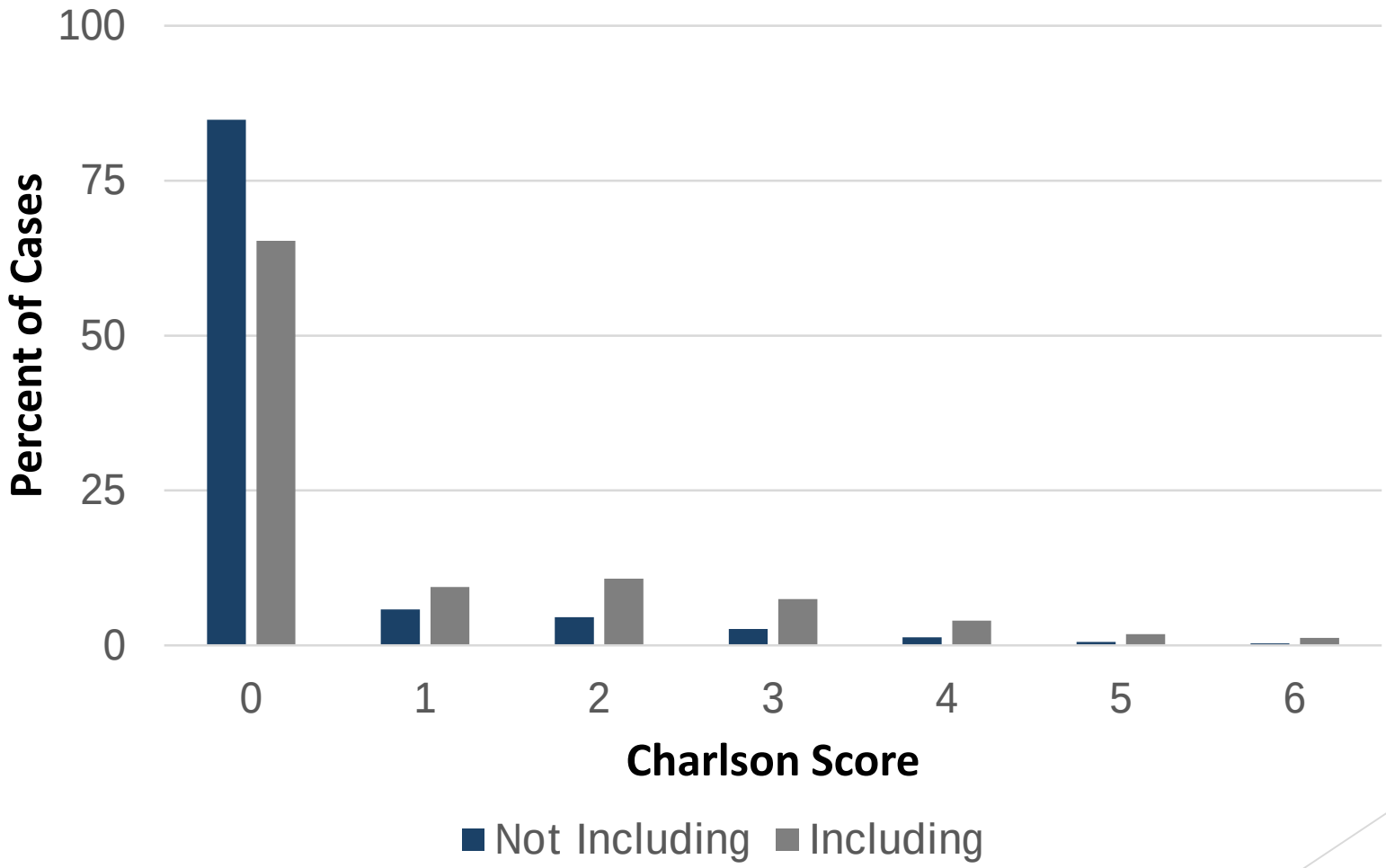
Charlson Comorbidity

- Developed in 1987 using medical records, multiple updates
- Summary measure derived from 19 specific medical conditions, each assigned a weight between 1 and 6
- Originally developed to predict 1 year all-cause mortality for hospitalized patients
- Algorithm available on the SEER-Medicare website
  - 2 versions (2000, 2014)

Charlson Comorbidity

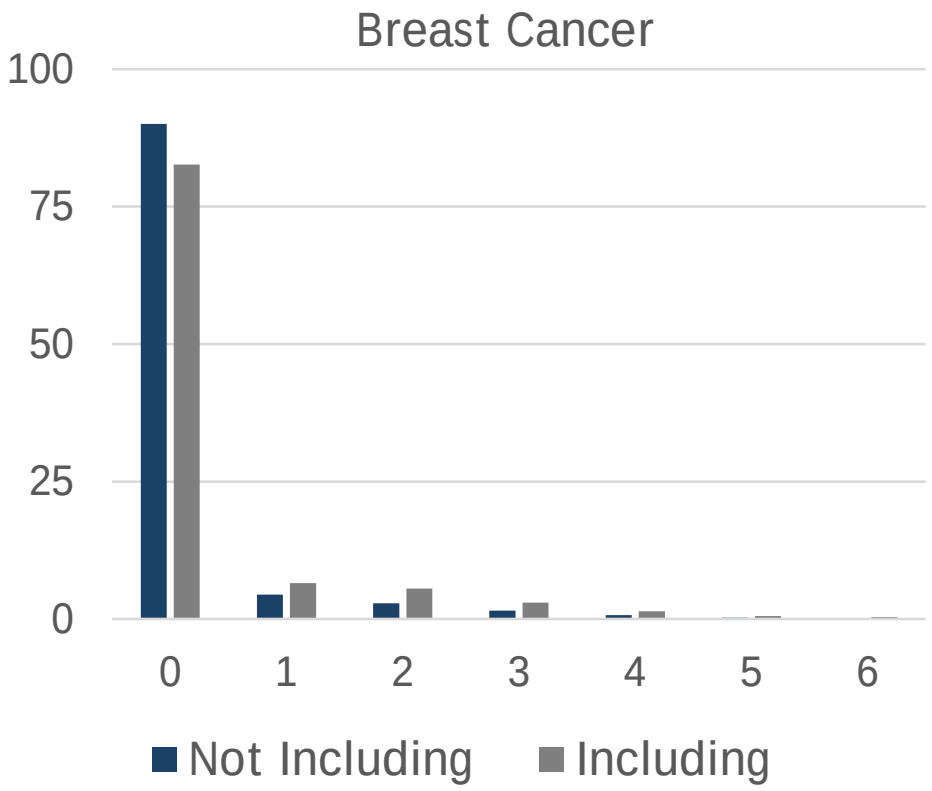
- Often uses a one-year look-back
  - Need to limit cohort to 66+ to capture 12 months
- May choose to not include month of diagnosis
- Theoretical range: 0 – 23
- Actual range: 0 – 9

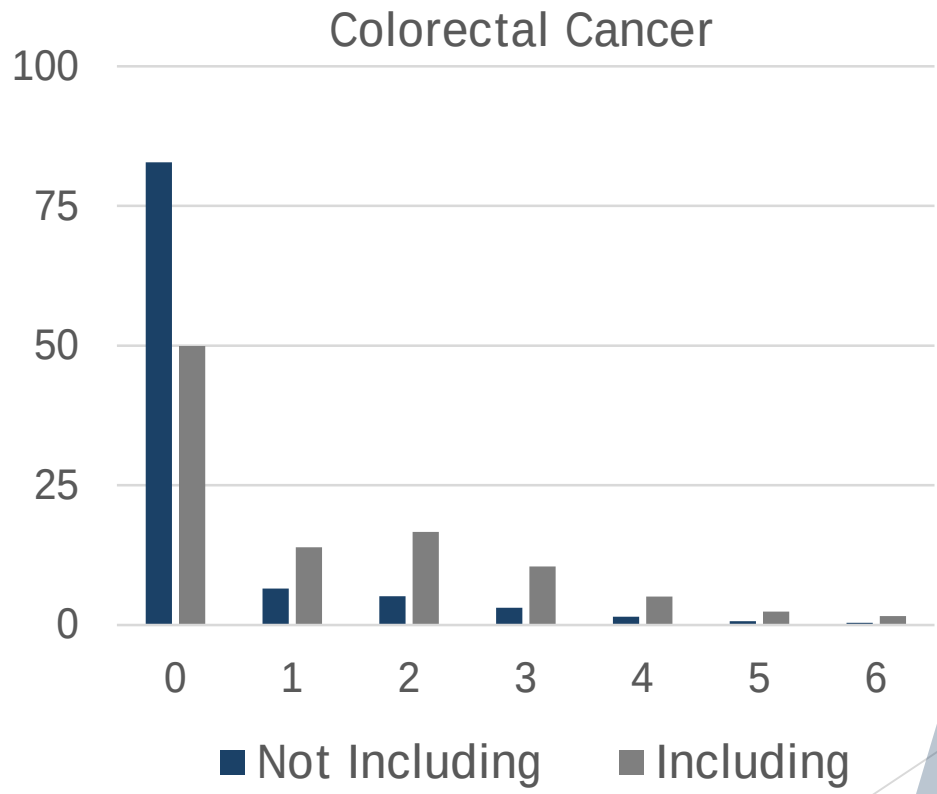
Consider Whether to Include Month of Diagnosis in Calculation



SCHOOL OF PUBLIC HEALTH
UNIVERSITY OF MINNESOTA

This Varies With Cancer Diagnosis





SCHOOL OF PUBLIC HEALTH
UNIVERSITY OF MINNESOTA

Charlson Comorbid Conditions

- Myocardial infarction
- Peripheral vascular disease
- Dementia
- Connective tissue disease
- Mild liver disease
- Moderate/severe liver disease (wt=3)
- Moderate/severe renal disease (wt=2)
- Any tumor (wt=2)
- Leuk/lymph (wt=2)
- Congestive heart failure
- Stroke
- COPD
- Ulcer disease
- Diabetes
- Diabetes with complications (wt=2)
- Paralysis (wt=2)
- AIDS (wt=6)
- Metastatic solid tumor (wt=6)
  - not included in SEER-Medicare algorithm

Charlson Comorbidity

- Generally in cancer patients, as comorbidity increases, treatment intensity decreases
  - Some treatment options may be contraindicated

Table 1 – Unweighted demographic characteristics of non-cancer control group and prostate cancer cohort stratified by treatment group

	Treatment group						
	Control	EBRT	BT	BT+EBRT	RP	RP+EBRT	Cryotherapy
Number	144 816	44 318	14 259	11 835	26 790	1557	2,115
Age at treatment (%)							
65-69	36.74	17.55	25.91	25.29	51.00	51.12	22.30
70-74	30.83	37.21	40.26	40.56	40.37	39.82	34.06
75-79	19.18	32.64	26.54	27.02	7.97	[b]	29.90
80-84	13.25	12.59	7.29	7.13	0.66	[b]	13.75
Race (%)							
White	83.33	83.10	88.70	84.36	86.70	84.59	82.95
Black	6.62	9.99	6.67	9.51	6.55	6.23	10.54
Hispanic	2.50	1.78	1.02	1.61	1.95	2.44	2.35
Asian	3.95	2.99	2.03	2.57	2.28	4.11	2.26
Other unknown	3.60	2.13	1.59	1.94	2.52	2.63	1.90
Median income (%)							
Q1	27.08	22.53	20.54	17.92	19.38	18.43	28.00
Q2	25.82	24.22	21.97	20.23	25.30	26.72	24.69
Q3	24.16	25.44	26.99	28.07	26.75	26.91	25.10
Q4	22.94	27.81	30.50	33.78	28.56	27.94	22.21
High school completion (%)							
Q1	27.36	23.50	21.01	20.81	18.30	20.36	27.23
Q2	25.54	25.05	23.70	22.61	22.55	21.77	24.88
Q3	24.58	25.83	26.32	26.13	26.47	27.17	25.64
Q4	22.52	25.63	28.97	30.44	32.69	30.70	22.25
Charlson score (%)							
0	69.11	62.11	65.57	62.92	68.28	65.19	54.50
1	19.68	24.36	23.68	25.56	23.45	25.05	29.13
2	6.85	8.43	7.21	7.72	5.98	7.19	10.36
3+	4.35	5.09	3.55	3.80	2.29	2.57	6.02
Clinical T stage (%)							
1	N/A	37.85	50.89	38.85	37.11	31.28	42.20
2x[a]	N/A	16.83	17.15	16.89	15.01	13.55	14.88
2y	N/A	11.32	5.91	13.38	10.40	12.14	9.14
2z	N/A	25.85	23.78	25.40	28.31	23.96	28.22
3 or 4	N/A	5.10	0.36	3.77	7.27	17.53	2.89
Unknown	N/A	3.05	1.91	1.71	1.90	1.54	2.67
WHO grade (%)							
1	N/A	4.92	3.45	2.18	4.45	2.12	1.09
2	N/A	57.81	80.01	52.49	63.85	40.98	52.92

Comorbidity Measures Rely on Reporting of Conditions on Health Claims

- Registries do not collect information on comorbidities, adding claims allows for measurement
- Weaknesses
  - Under-diagnosis
  - Under-reported on claims
    - Ex: Hypertension, osteoporosis and diabetes
- Conditions may not be recognized or reported until they lead to problems
  - Ex: osteoporosis first noted on a fracture claim

Inference is Affected by Timing of Claims Report

- People with poorer bone density may be more likely to be treated with bisphosphonates
  - This might lead to the conclusion that bisphosphonates cause fractures
- Dementia/cognitive impairment may influence decisions even if they aren't formally diagnosed or coded
  - How would that show up in a study of dementia incidence?
  - Does it make sense that chemotherapy 'protects against dementia'?

Cancer Characteristics

- May affect treatment recommendations and use

Table 1 – Unweighted demographic characteristics of non-cancer control group and prostate cancer cohort stratified by treatment group

	Treatment group						
	Control	EBRT	BT	BT+EBRT	RP	RP+EBRT	Cryotherapy
Number	144 816	44 318	14 259	11 835	26 790	1557	2,115
Age at treatment (%)							
65-69	36.74	17.55	25.91	25.29	51.00	51.12	22.30
70-74	30.83	37.21	40.26	40.56	40.37	39.82	34.06
75-79	19.18	32.64	26.54	27.02	7.97	[b]	29.90
80-84	13.25	12.59	7.29	7.13	0.66	[b]	13.75
Race (%)							
White	83.33	83.10	88.70	84.36	86.70	84.59	82.95
Black	6.62	9.99	6.67	9.51	6.55	6.23	10.54
Hispanic	2.50	1.78	1.02	1.61	1.95	2.44	2.35
Asian	3.95	2.99	2.03	2.57	2.28	4.11	2.26
Other unknown	3.60	2.13	1.59	1.94	2.52	2.63	1.90
Median income (%)							
Q1	27.08	22.53	20.54	17.92	19.38	18.43	28.00
Q2	25.82	24.22	21.97	20.23	25.30	26.72	24.69
Q3	24.16	25.44	26.99	28.07	26.75	26.91	25.10
Q4	22.94	27.81	30.50	33.78	28.56	27.94	22.21
High school completion (%)							
Q1	27.36	23.50	21.01	20.81	18.30	20.36	27.23
Q2	25.54	25.05	23.70	22.61	22.55	21.77	24.88
Q3	24.58	25.83	26.32	26.13	26.47	27.17	25.64
Q4	22.52	25.63	28.97	30.44	32.69	30.70	22.25
Charlson score (%)							
0	69.11	62.11	65.57	62.92	68.28	65.19	54.50
1	19.68	24.36	23.68	25.56	23.45	25.05	29.13
2	6.85	8.43	7.21	7.72	5.98	7.19	10.36
3+	4.35	5.09	3.55	3.80	2.29	2.57	6.02
Clinical T stage (%)							
1	N/A	37.85	50.89	38.85	37.11	31.28	42.20
2x[a]	N/A	16.83	17.15	16.89	15.01	13.55	14.88
2y	N/A	11.32	5.91	13.38	10.40	12.14	9.14
2z	N/A	25.85	23.78	25.40	28.31	23.96	28.22
3 or 4	N/A	5.10	0.36	3.77	7.27	17.53	2.89
Unknown	N/A	3.05	1.91	1.71	1.90	1.54	2.67
WHO grade (%)							
1	N/A	4.92	3.45	2.18	4.45	2.12	1.09
2	N/A	57.81	80.01	52.49	63.85	40.98	52.92

Cancer-Specific Measures

- Cancer stage/extent of disease
  - Local, regional, distant
- Nodes
  - Positive vs. testing
- Grade

- Not available from SEER registries
  - Genetic characteristics
  - Family history
  - Surgical margins
  - Behavioral risk factors

- Site specific variables
  - Tumor size
  - Extension into surrounding tissues
  - Nodal involvement
  - Ex:
    - BRCA
    - Gleason
    - PSA Values

Behavioral Risk Factors and Family History in Claims

- Smoking
- Obesity
- Alcohol use
- Family history of cancer

- ICD-9/10 diagnosis codes exist, but:
  - No lung cancer cases with a smoking code?
  - Obesity code is required for bariatric surgery, but rarely coded otherwise.
- Question: is the risk factor essential to your study?

Summary—Covariates

- Where do you find the information?
  - Demographic
    - PEDSF
  - Clinical
    - Claims/PEDSF
- *Can* you find the information?
  - Lifestyle/Behavioral Risk Factors

Questions?

SCHOOL OF PUBLIC HEALTH
UNIVERSITY OF MINNESOTA

Measuring Treatment

Segment 4

SEER-Medicare Training 2019

Helen Parsons, PhD, MPH

SCHOOL OF PUBLIC HEALTH
UNIVERSITY OF MINNESOTA

Common Cancer Directed Treatments Include

- Surgery
- Radiation
- Chemotherapy/Biologics
- Combinations

Thinking about Measuring Cancer-Directed Therapy

- SEER and Medicare both measure cancer directed therapies, but they measure them differently

- Understanding the similarities and differences is essential to deciding which source/measure to use and determining what to do if the two measures point in different directions

Why Measure Treatment?

- Who gets what?
  - May vary by region, or based on demographics
- Does procedure influence outcome?
  - Can assess in the real world vs. trials
- Assess quality of treatment
  - Full range of care settings
- Are there long-term consequences?
- Evaluate determinants of outcomes
- Evaluate determinants of disparities

Measuring Surgery

Where Can Surgery be Measured?

Big Picture

- Both SEER and Medicare sources may be used
- SEER data – predefined list of procedures, no dates. Only the 'biggest' operation noted
  - Easier but limited
- Medicare data – detailed information about all procedures. Requires creating definitions and windows based on coding and billing rules for Medicare
  - more complete but more difficult to use

Measuring Surgery in SEER (I)

First course of treatment:

- Definition: “First course of treatment includes all methods of treatment recorded in the treatment plan and administered to the patient before disease progression, recurrence or treatment failure”
- Since 1998, includes all planned surgery, or surgery within 1 year if no evidence of recurrence or failure
- Consult the SEER Program Coding and Staging Manual for more details

Measuring Surgery in SEER (II)

What is coded?

- SEER codes most extensive procedure
  - Follows established hierarchy
  - Biopsy → Lumpectomy → Mastectomy
    - Coded as Mastectomy
  - Cannot determine course of events
- No information about provider available
  - Important if evaluating processes of care

Measuring Surgery in SEER (III)

What information is available?

- Separate variables for surgery of primary site, lymph node evaluation, distant sites, and reconstruction for all organs
- Site Specific Surgery Information for most organs
  - Examples include: oral cavity, parotid gland, pharynx, esophagus, stomach, colorectal, anus, liver and bile ducts, pancreas, larynx, lung, bones and soft tissues, spleen and lymph nodes, skin, breast, cervix, uterus, ovary, prostate, testis, kidney, bladder, brain, thyroid
  - Note: Surgery codes may have different meaning by cancer type
- Coding changes in 2004 to be consistent with collaborative staging (i.e., separate variables pre- vs. post 2004 to evaluate treatments)
- Check SEER coding manuals for specific codes as there have been many recent changes, fully expect continued updates

Measuring Surgery in SEER (IV)

Lymph Node Evaluation

Can Determine:

- Were nodes examined? How many?
- Were nodes positive? How many?
- Note: Variables available for certain years and certain cancers only (e.g., colon cancer)
- This information is NOT available in Medicare other than some information on whether specific procedures were performed (not results)

Caveat Emptor

- SEER coding changes must be kept in mind when evaluating time trends
- Major changes in trends around the time coding changes occurred must be scrutinized
  - Example: Check before and after 1998 and 2003 as there were major changes in how staging/treatment were reported

Measuring Surgery in Medicare (I)

- Depending on the measures of interest, may need to use multiple file types within Medicare data to identify surgical measures
- At a minimum, individuals hospitalized for cancer directed surgery will produce two bills
  - Hospital bill covering the room, nursing services, etc. (MedPAR File)
  - Physician bill for performing the surgery (NCH/Carrier Claims)
- Cancer-directed surgery in an outpatient setting will have minimum of one bill
  - Outpatient facility bill covering the physician performing the surgery, surgical suite, nursing services, etc. (Outpatient File)
- Medicare reimburses for these services using distinct billing codes and fee schedules depending on the setting (e.g., inpatient vs. outpatient)

Measuring Surgery in Medicare (II)

Hospital Bills (MedPAR File)

- Hospitals bill for surgical services provided using ICD-9 procedure codes (ICD-10 after October 2015)
- Each hospitalization allows coding of up to 25 procedures
- Each procedure will have an associated procedure date

Outpatient Facility/Physician Bills

- Outpatient and NCH/Carrier claims code for services using CPT/HCPCS codes
- In general, most services are billed in subcomponents (e.g., surgery, anesthesiology, pathology)
- Includes procedure dates as well as the performing surgeon identifier

Note: ICD-9 and CPT codes have different levels of detail for understanding the surgery performed

Measuring Surgery in Medicare (III)

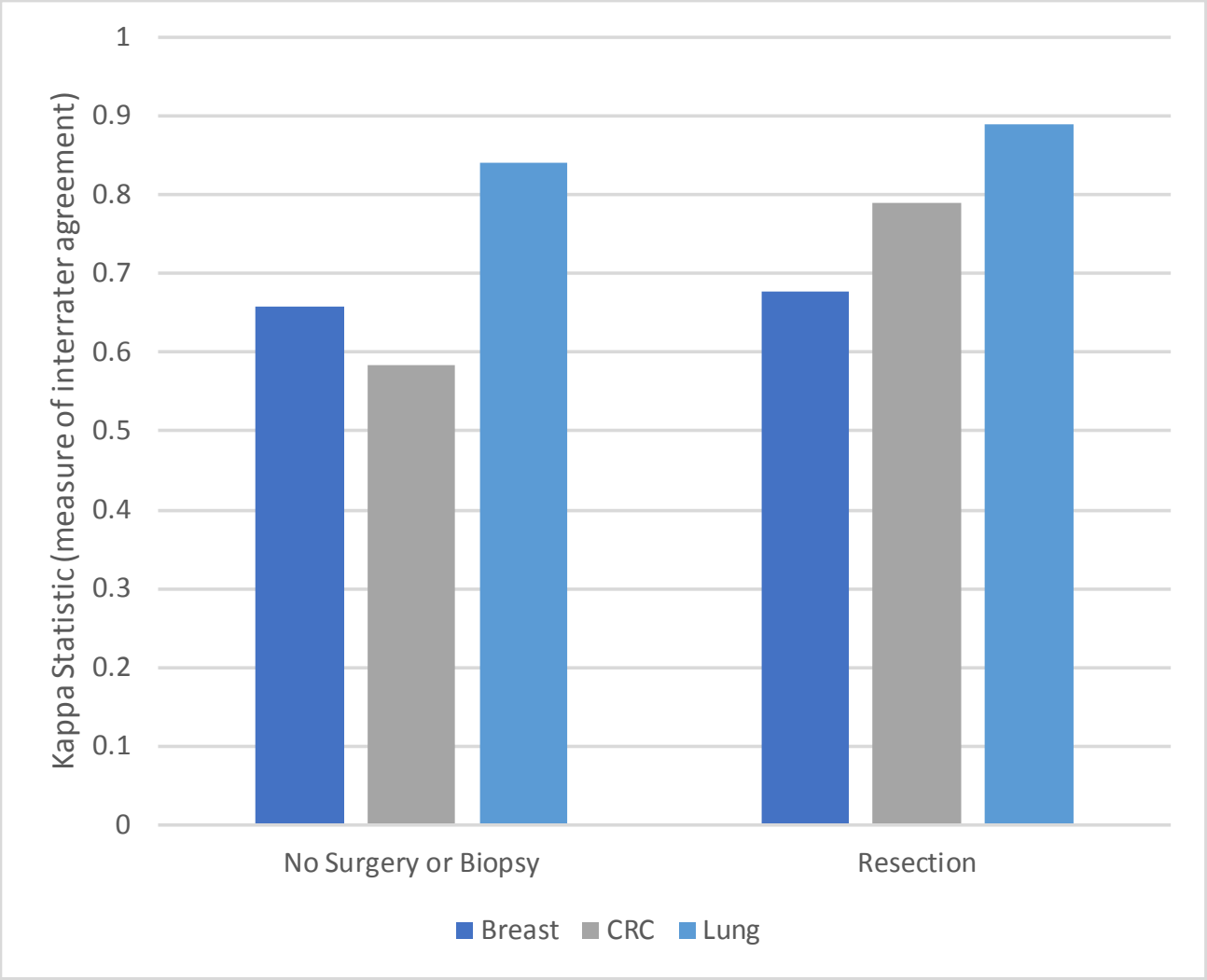
- Bundling refers to paying **surgeons** a global payment for 'routine care' related to a procedure
  - Pre-op and post-op care
    - Cannot tell whether a follow-up visit occurred or if only minor problems happened
    - Major problems are not included in bundles
    - Bundling rules have changed over time
  - Bundling does not cross providers
  - Anesthesiologists, pathologists, imaging, etc. are not subject to bundling

There is Not an Exact Match Between ICD and CPT Coding Systems

	ICD-9* (MedPAR)	CPT (NCH/Carrier and Outpatient)
Lymph node evaluation	**40.23** Excision of axillary lymph node **40.3** Regional lymph node excision	**38500** biopsy or excision of lymph node(s)—superficial 38525 Deep axillary node(s) **38740** axillary lymphadenectomy superficial **38745** axillary lymphadenectomy, complete **38792** Injection procedure, lymphangiography for identification of sentinel lymph node

*ICD-10 after October 1st, 2015

Agreement Between SEER and Medicare About Surgery Can Vary Significantly by Cancer Type and Procedure



*Includes breast, colon, lung cancers diagnosed 2011-2015 who were 'likely to have complete claims' within 4 months of diagnosis using combined MedPAR, physician, and hospital outpatient files

Measuring Surgery in SEER vs. Medicare

Key Take-Aways

- Both sources provide information on cancer-directed surgery
- SEER provides high-level first course of treatment information
  - General, site-specific surgical codes
  - Easier to create and measure
  - Doesn't include physician/provider information or procedure dates
- Medicare claims include more detail:
  - Information on dates of surgery, performing surgeon
  - Can identify more detailed information on the procedure using ICD/CPT codes
  - Windows defined by researcher and can be study-specific

Measuring Radiation Therapy

Where Can Radiation Therapy be Measured?

Big Picture

- As with cancer-directed surgery, both SEER and Medicare sources may be used
- SEER data – predefined list of radiation therapy categories, sequence with surgery, but no dates. Only includes therapy provided as the "first course of treatment"
  - Easier but limited
- Medicare data – detailed information about all procedures. Requires creating definitions and windows based on coding and billing rules for Medicare
  - More complete, with lots of detail on dates of service, dosing, etc.
  - Claims for radiation therapy will typically be found in outpatient or NCH/Carrier claims

Radiation Therapy in SEER (I)

- First course of cancer-directed therapy
- Limited to treatment in the first 4 months or 12 months depending on year of diagnosis
- Coded as:
  - No radiation
  - Radiation used (5 categories)
  - Radiation refused
  - Recommended but unknown if used
  - Missing
- No information on dose
- No information on intended target (primary tumor or secondary spread)

Radiation Therapy in SEER (II)

4 Categories of Radiation Reported

- Beam
- Radioactive Implants
- Radioisotopes
- Combination of Beam with one of the others
- Also,
  - Radiation, type not specified
  - Other

Radiation Therapy in SEER (III)

Timing of Radiation Reported by SEER

- Radiation prior to surgery
- Radiation after surgery
- Radiation before and after surgery
- Intraoperative radiation
- Intraoperative radiation with other radiation
- Sequence unknown but both surgery and radiation were given

Medicare Information on Radiation Therapy (I)

- Radiation is billed as a procedure, allowing for evaluation of dates of service and dosing
- Can differentiate and find combinations of:
  - External beam RT
  - Brachytherapy
  - Intensity Modulated Radiation Therapy (IMRT)
  - Proton beam RT

Medicare Information on Radiation Therapy (II)

How do you find Radiation Therapy?

- MedPAR:
  - ICD-9 procedure codes: 92.21-92.29 (ICD-10 after Oct 1, 2015)
- Outpatient and NCH/Carrier Files (Most bills found here)
    - Initial imaging
    - Planning
      - Examples: 77370 (CPT) – Radiation Physics Consultation, 77300 (CPT) – Basic Dosimetry
    - Actual Treatment Delivery
      - Examples: 77402 (CPT) - Radiation treatment delivery, single treatment area, single port or parallel opposed ports, simple blocks or no blocks
    - Oversight of Treatment
      - Examples: 77431 (CPT) - Radiation therapy management

Medicare Information on Radiation Therapy (III)

- CPT codes allow for understanding RT with the most detail; charges occur at all points in RT
- As with other services, codes are added or discontinued over time so evaluate your list carefully

Medicare Information on Radiation Therapy (IV)

Creating Measures of Radiation Therapy Use

- RT is tricky because it spans several phases (planning, monitoring and treatment) so multiple codes, units, dates should be observed.
- Can build algorithms to categorize:
  - Type(s) of RT
  - RT consult or planning only but no TX
  - Incomplete TX (how many doses would that be?)
  - RT completed (how many doses would that be?)

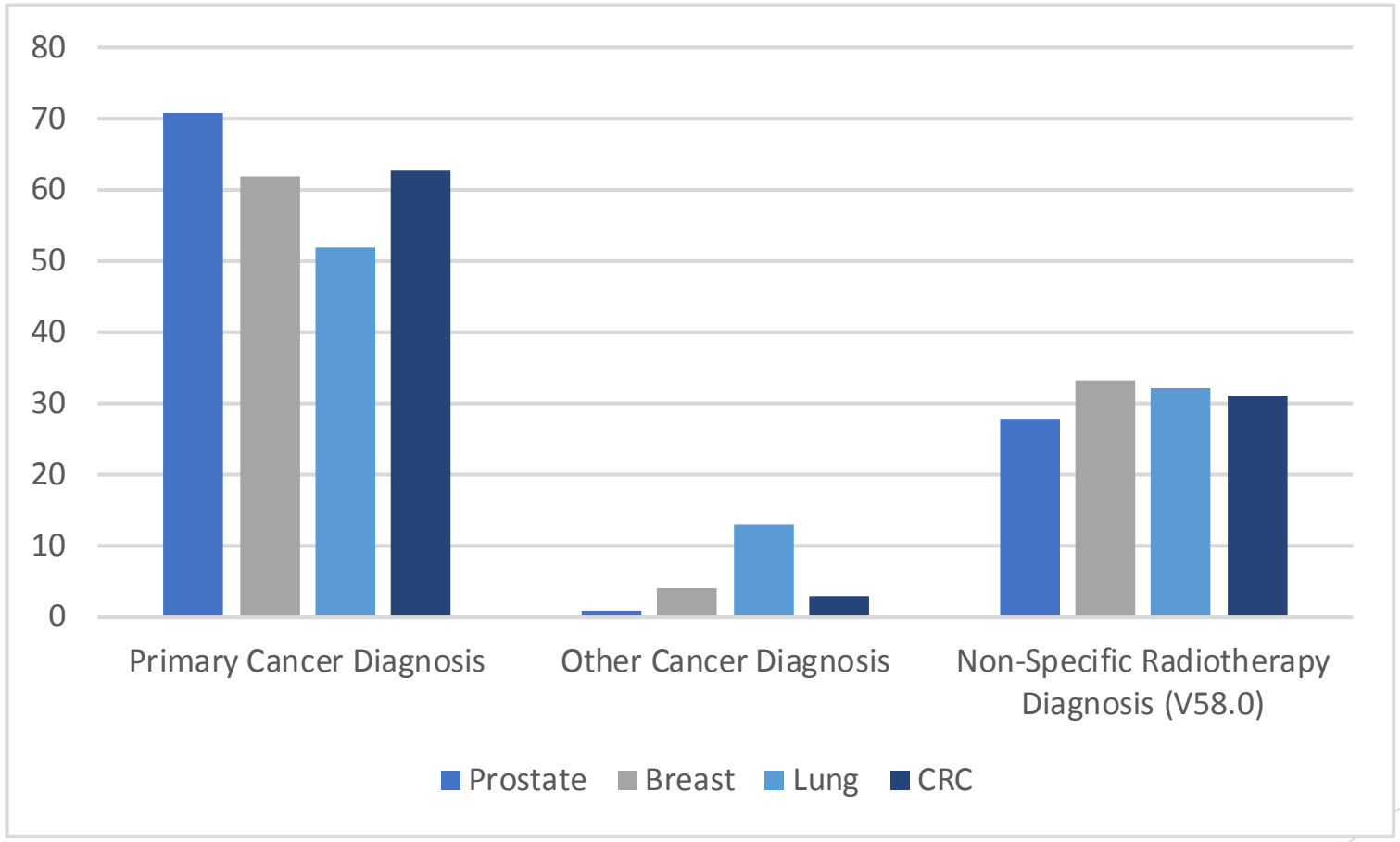
Medicare Information on Radiation Therapy (V)

- Measuring Radiation Days/ # of Services may require use of M/T/U/S variable in the NCH/Carrier file and revenue center count in the outpatient file
- Radiation is typically administered daily over a fixed period of time
  - E.g., 5 days a week for 6 weeks
  - This may be reflected in claims a number of ways
  - If you are attempting to count RT services, you will need to be extremely careful about what you are counting…
    - One bill/line item for each day of service (30+ line items with M/T/U/S=1)
    - One bill per week with 5 treatments for 6 weeks (6 bills with M/T/U/S=5)
    - One bill for a month with 20 treatments and another for 2 weeks of treatments
    - **Plus, there will be additional claims for the radiologist who is overseeing the RT, treatment planning, port films, etc.**

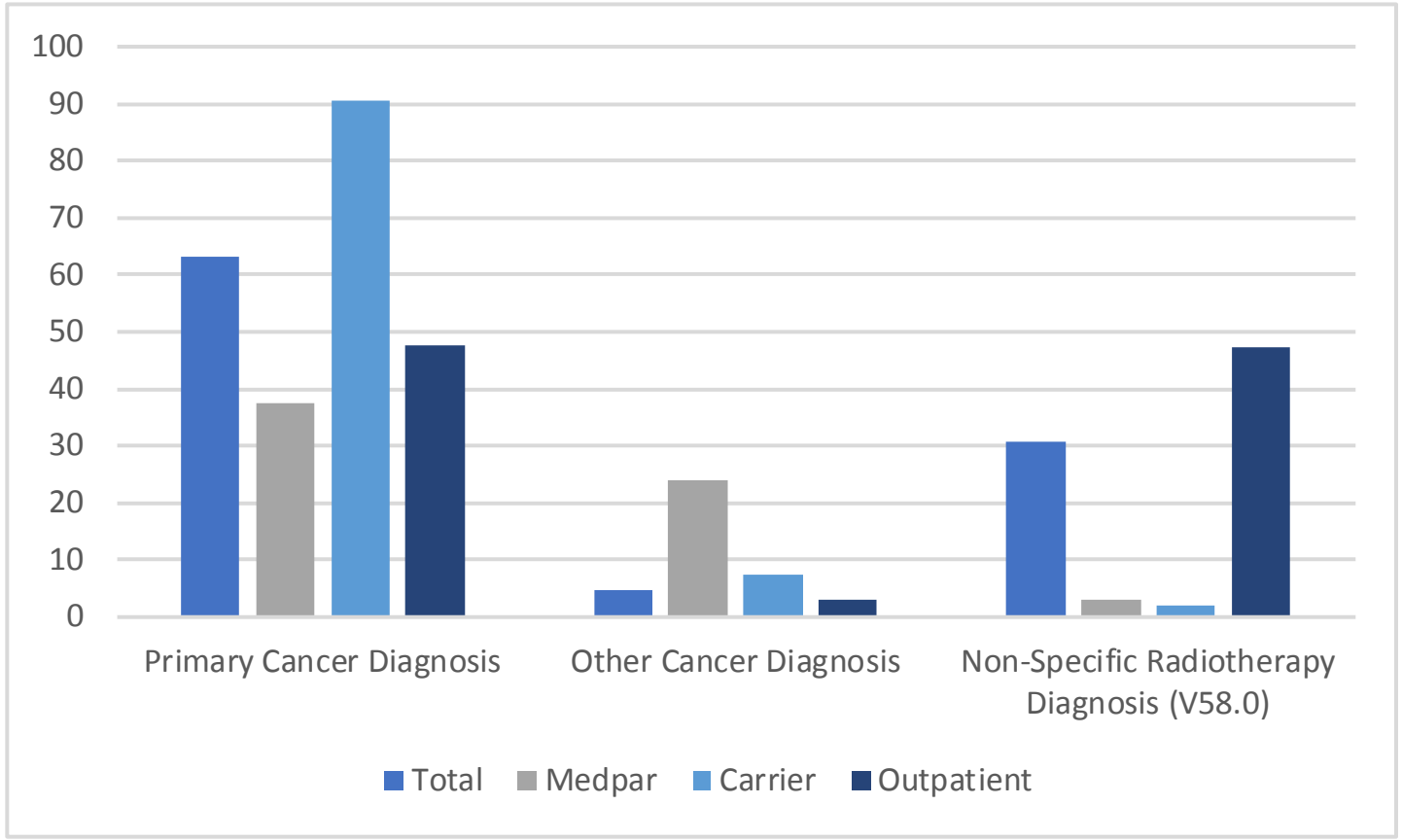
Diagnosis Codes Associated with Radiation Procedures

- Should it be assumed that the receipt of radiation is targeted to the incident cancer?
  - – If not, this has implications for concordance between SEER and Medicare and could maybe be a way to study disease spread and/or palliative care

ICD-9 Diagnosis Codes Associated with Radiation Procedure Codes in Claims



ICD-9 Diagnosis Codes Associated with Radiation Procedure Codes in Claims



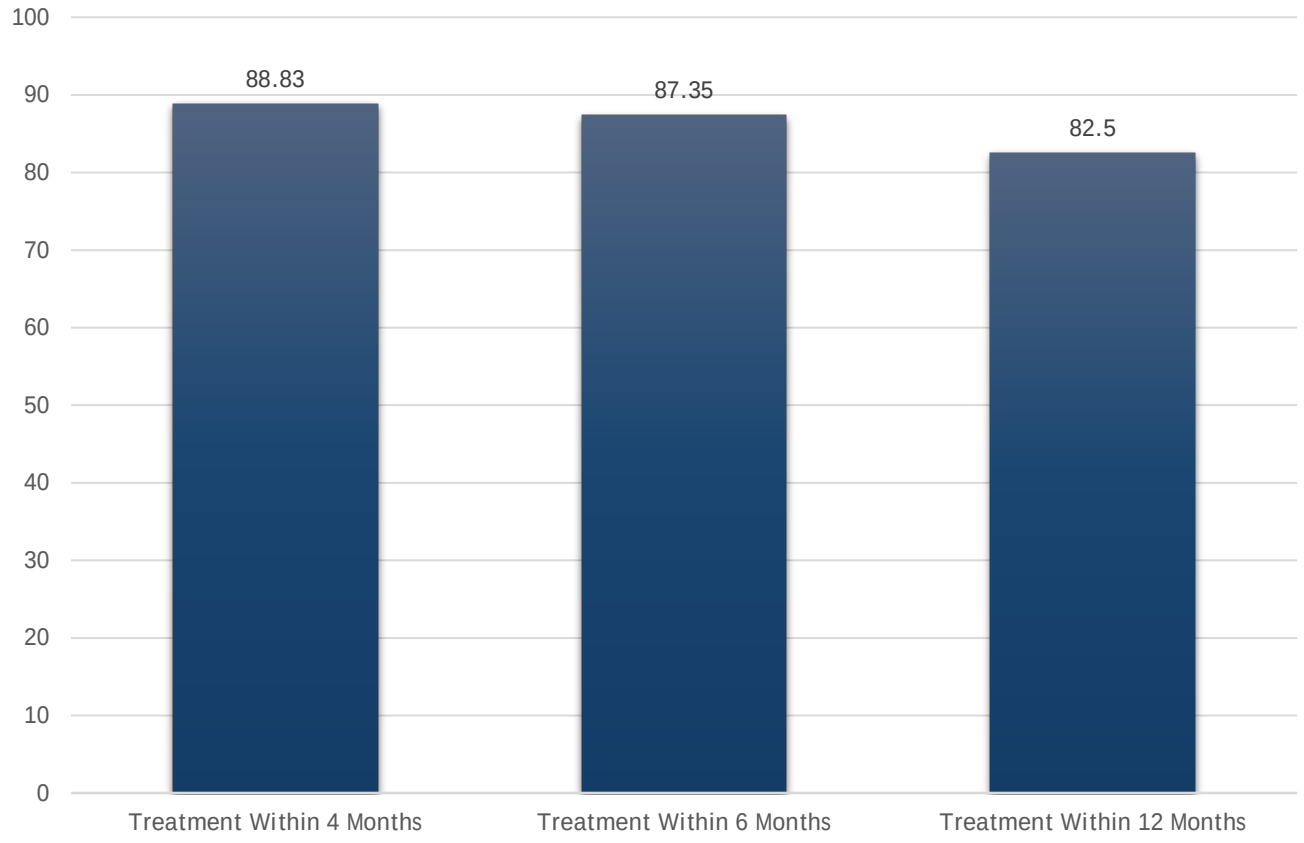
Concordance between SEER and Medicare on Radiation Therapy

Example

- Breast cancer diagnosed in 2014
- Restrict to first and only cancer diagnosis
- Treatment likely to be found in claims data:
  - Enrolled in Medicare Part A and Part B at least one month in 2014
  - No HMO enrollment in 2014 or 2015
- Search claims for evidence of radiotherapy using list of ICD/CPTs:
  - MedPAR
  - NCH/Carrier
  - Outpatient
- N = 12,150
- Create a cohort of individuals “likely to have complete claims”

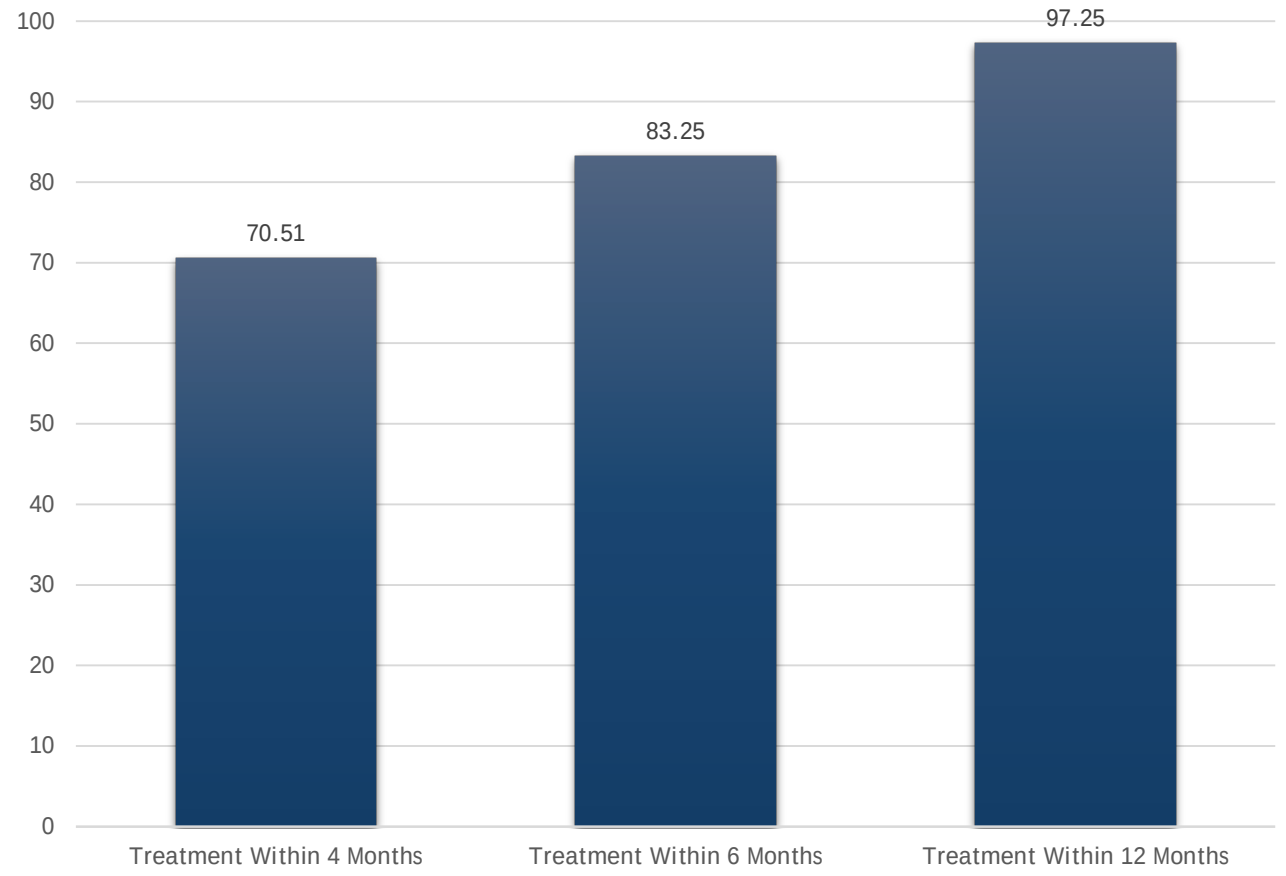
With Medicare as the Gold Standard

- We are able to find SEER-based evidence of RT for 84% of the people who have Medicare claims for RT



With SEER as the Gold Standard

- We are able to find Medicare-based evidence of RT for 90% of the people who SEER said received RT



Why Differences Between SEER and Medicare?

- More radiation reported in Medicare sources may be due to factors including:
  - SEER fails to ascertain the RT (outpatient treatments are tougher to find)
  - mismatches between treatment windows under Medicare and SEER
  - radiation aimed at secondary cancer sites
- Less radiation reported in Medicare than SEER may be because treatment received from non-Medicare providers (VA, I.H.S., private insurance)
- Limiting analysis to people likely to have complete claims and with cancer diagnosis in claims may help concordance

Radiation Therapy- Problems and Pitfalls with Measurement

- Intent cannot be measured directly
  - Curative vs. palliative vs. ???
- Cannot identify role of patient choice
- In some cases, SEER notes "treatment offered but not accepted" but likely incomplete
- Most important information regarding quality not available from either SEER or Medicare
  - Technical and pathologic details beyond the scope of collected data
  - Quality must be measured indirectly

Measuring Radiation Therapy in SEER vs. Medicare

Key Take-Aways

- Both sources provide information on radiation therapy
- There is generally high agreement between SEER and Medicare with regard to radiation therapy after diagnosis of cancer
- SEER provides high-level first course of treatment information
  - General radiation therapy codes
  - Easier to create measures
  - Doesn't include physician/provider information or procedure dates
- Medicare claims require knowledge of billing and coding rules to identify how and where bills for radiation therapy will appear
  - Allows for more detailed information on dates of therapy, providers
  - Allows for evaluation of therapy timing, dosing, payments

Measuring Chemotherapy/Biologics

Where Can Chemotherapy be Measured?

Big Picture

- Chemotherapy information from SEER is not available in the SEER-Medicare data because of concerns about under-ascertainment.

- Medicare part B pays for chemotherapy administration and for specific drugs given by injection/IV and their equivalents or metabolites.
- Medicare part D pays for all other oral medications without an IV equivalent.
  - Part D data available from 2007 forward

Caveats with “Chemotherapy”

- Use term “chemotherapy” to broadly include all medications aimed at treating/controlling cancer
- Some agents will have non-cancer-treatment indications:
  - Other conditions (e.g., methotrexate)
  - Cancer prevention (e.g., tamoxifen)

Identifying Chemotherapy in Medicare Claims

- Chemotherapy can be found in multiple files based on CMS policy
- IV chemotherapy is typically found in the Outpatient and NCH/Carrier files
  - Expect codes for both:
    - Agent (many but not all start with a J)
    - Delivery: 964XX, 965XX, Q0083-Q0085
- Oral chemotherapy drugs could be found in either Outpatient/NCH(Carrier) or Part D depending on whether it has an IV equivalent
- There are challenges….
  - Inpatient Chemotherapy is challenging because there are no CPT codes in the MedPAR file (i.e., cannot identify agent)
  - ICD diagnosis codes and revenue center codes that apply to chemotherapy are non-specific

Where will I find Chemotherapy Bills in Medicare Claims?

Example 1: IV Drugs or those with IV Equivalent (Part B Drug)

	MedPAR	NCH/Carrier	Outpatient	Part D (PDE)
Fluorouracil (IV)	*	CPT	CPT	
Capecitabine (Oral equivalent)	*	CPT	CPT	

Example 2: Drugs with no IV Equivalent (Part D Drug)

	MedPAR	NCH/Carrier	Outpatient	Part D (PDE)
Gefitinib	*			NDC

* Inpatient settings only report general chemotherapy administration(i.e., don't know exact drug)

Other Services/Codes in the Medicare Claims Related to Chemotherapy Use

- Equipment- Pumps/Reservoirs/Ports
- Prehydration/Premedication
- Anti-emetics - found in the HCPCS/CPT. Medicare pays for anti-emetic drugs given within 48 hours of chemotherapy if the medication is given IV or orally as an alternative to IV.
- Antiemetics outside this time range may be found in Part D data
- CMS issues coverage decisions on occasion such as:
  - Only for selected chemotherapies
  - Only when given with other medications

Chemotherapy in Part D

- Chemotherapies that are only available in oral form (no IV equivalent or metabolite) will be found in Part D data
- Not all enrollees will have Part D. Expect loss of sample size due to this requirement
- Part D data contain information on:
  - Drug
  - Drug type
  - Dosage
  - Date the prescription was filled
  - Copayments and deductibles
- But not:
  - Prescribing MD
  - Date prescription was written
  - Diagnosis

As With Radiation Therapy, You Will Need to Be Extremely Careful When Considering Measurement of Dosing or Regimen

- May see:

  Outpatient/NCH(Carrier)

  - One bill/line item for each day of service (30+ line items with M/T/U/S=1 or revenue center count=1)
  - One bill per week with 5 treatments for 6 weeks (6 bills with M/T/U/S=5 or revenue center count=1)
  - Note: CPT code provides implied dosing information (e.g., CPT J9190 fluorouracil, 500mg)

  Part D

  - One bill per prescription fill (e.g., NDC 0310-0482-30, Gefitinib 30 tablets in 1 bottle)

Caveat: Not All Dosing May Fit Within Clinical Norms or Practices; Will Need to Decide How to Classify These Individuals

Example: Colon Cancer

	Eligible cases	% with any claim for Irinotecan	Median daily dose (mg)	Daily dose range (mg)	% with a single daily dose >800 mg
Stage 1	9,581	0.2%	400	160-6,820	66.7%
Stage 2	12,495	1.2%	500	20-18,920	61.9%
Stage 3	9,105	6.7%	520	20-12,180	56.3%
Stage 4	5,870	17.3%	240	20-16,480	48.0%

Measuring Chemotherapy in SEER-Medicare

Key Take-Aways

- Chemotherapy information not available from SEER
- Must use Medicare claims to identify chemotherapy treatment
- With the exception of chemotherapy administered in an inpatient (hospital) setting, you can determine the drug administered, dates of therapy, and providers

Key Take-Aways (I)

Why Might SEER/Medicare Disagree on Cancer Directed Treatment?

- SEER may find care that is not covered by the Medicare program—VA, Indian Health Service, Private Insurance
- SEER will have a harder time finding cancer care that is delivered out of the SEER area, particularly if there is no record of the care by a local provider
- Medicare might find care that isn't considered part of the first course of treatment

Key Take-Aways (II)

- SEER and Medicare both measure some aspects of cancer treatment. Differences between the two sources are to be expected.
- SEER coding of treatments varies by cancer type and over time
- Medicare coverage of cancer therapy is defined by policy. In general, multiple claim types will be needed to completely ascertain services received.

Summary: SEER vs Medicare in Measuring Cancer Directed Therapy

	SEER	Medicare
Surgery	First course of treatment, most “major” surgery in a category	ICD-9 (inpatient) and CPT codes (all other sources) for specific services provided; dates, treatment windows can be created
Radiation	Limited information—type	ICD-9 (inpatient) and CPT codes (all other sources) for specific services provided; dates, treatment windows can be created
Chemotherapy	Not Available	Policy-based rules determine whether to expect to find use in NCH/OP/DME/HHA? vs. Part D files
Combination	For some cancers, sequence of surgery and radiation noted	Use dates to determine timing

Final Thoughts

- SEER-Medicare allows for identification of cancer-directed treatment including surgery, chemotherapy and radiation
- No gold standard for identifying therapy between SEER and Medicare
  - Mismatches expected due to scope of treatments identified and treatment definitions between sources
- Different sources of care (e.g., inpatient, outpatient) have different coding conventions (e.g., ICD vs. CPT) and will be found in different files
- Consider partnering with clinical experts to ensure treatment definitions are clinically valid (e.g. treatment windows, included codes)

Questions?

Lunch

We will return at 2:15 ET / 1:15 CT

Measuring Outcomes

Segment 5

SEER-Medicare Training 2019

Stephanie Jarosek, PhD

Outcomes

- After cancer diagnosis and treatment, a variety of outcomes may be of interest
  - Short term complications/adverse events
  - Long term complications/sequelae
  - Recurrence
  - Survival/Death
  - Cost of care

The Opportunity

- SEER Registries have information after diagnosis but do not conduct longitudinal follow-up of patients
  - Only vital status is linked after registration period
- The Medicare data are longitudinal and have the potential to capture outcomes.
  - BUT these are claims, not clinical records

Complications and Adverse Events

Complications/Adverse Events

- Cancer treatments can result in complications
  - Acute—arise quickly during or immediately after treatment
  - Chronic—persist after initial cancer treatment

Acute Adverse Events Following Surgery

- For inpatient procedures, acute adverse events related to surgery may be identified from diagnoses and procedures on the hospital claims.
  - Caveat: Some of these “complications” will actually be pre-existing conditions
- With shorter lengths of stay, many post-surgical complications will occur outside the hospital.
  - Caveat: Surgeons are paid on a global basis, which includes all routine post-operative services. Few post-operative complications not requiring rehospitalization will appear in the claims

In-Hospital Complications

Rueth et al. J Thorac Cardiovasc Surg 2012; 143: 1314-23

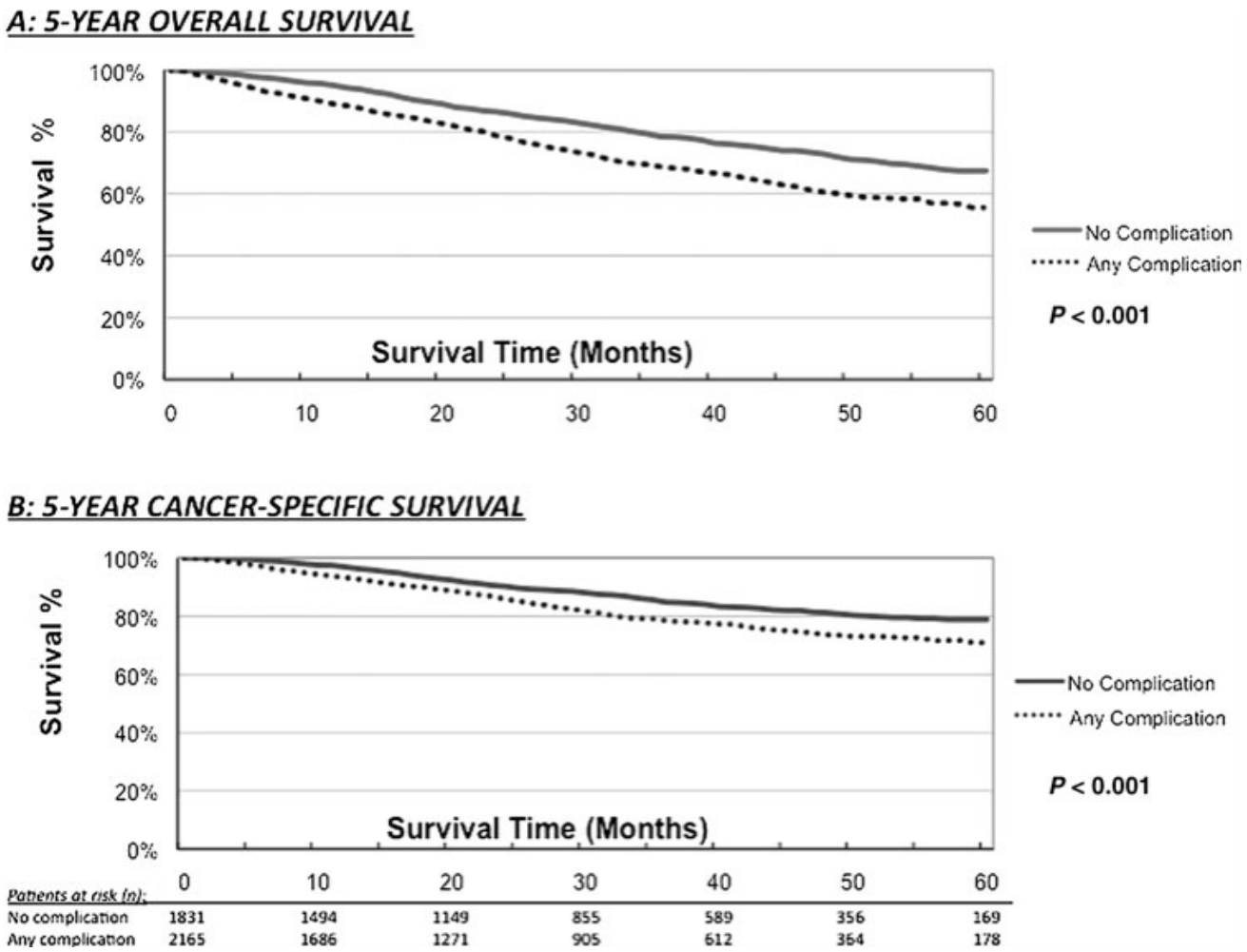
TABLE 3. Pulmonary, cardiac, and noncardiopulmonary complications (major complications separated by category)

	Complication occurrences identified (no.)	Population with complication (%)
Pulmonary*	2174	38.3%
Pneumothorax	705	16.9%
Atelectasis	525	12.6%
Pulmonary Insufficiency	253	6.1%
Respiratory failure	177	4.2%
Pneumonia	127	3.1%
Other	387	9.2%
Cardiac†	1,086	24.5%
Atrial fibrillation	946	22.7%
Acute myocardial infarction	56	1.3%
Cardiac arrest	28	0.7%
Cerebrovascular accident	14	0.3%
Other	42	1.0%
Noncardiopulmonary‡	837	16.1%
Wound or surgical	300	7.2%
Nonpneumonia infectious disease	216	5.2%
Hematologic	208	4.9%
Renal or hepatic	113	2.7%

*In total, 1598 patients had 2174 pulmonary complications. †In total, 1020 patients had 1086 cardiac complications. ‡In total, 674 patients had 837 noncardiopulmonary complications.

Long-term Impact of Surgical Complications

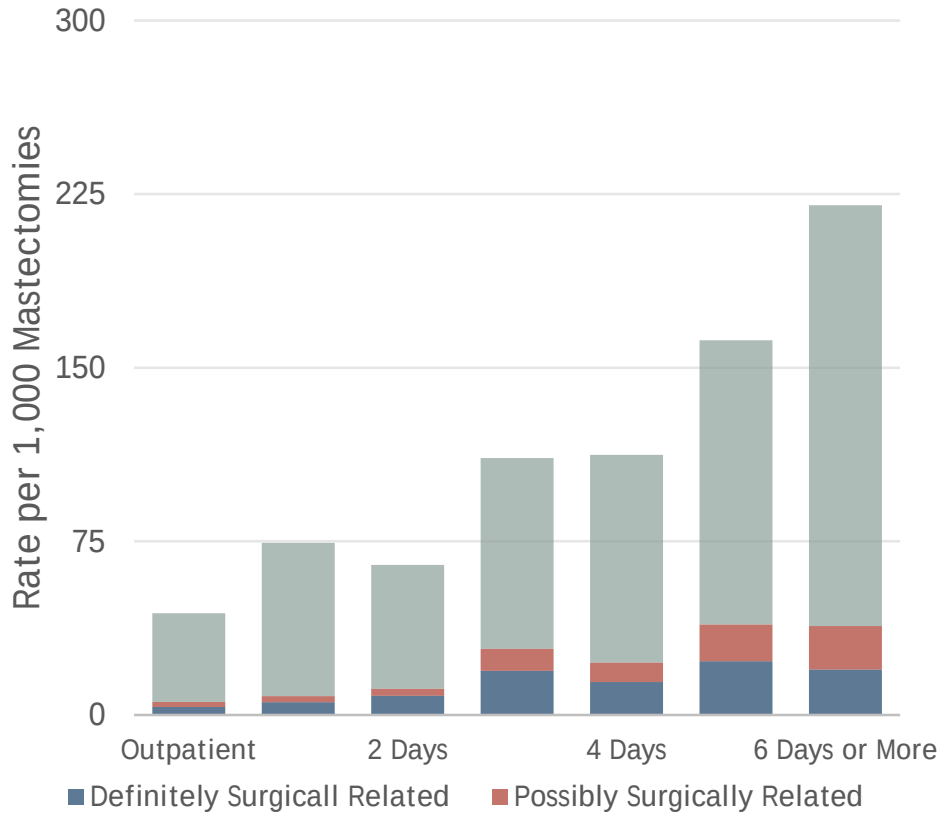
Rueth et al. Ann Surg 2011; 254: 368-74



SCHOOL OF PUBLIC HEALTH
UNIVERSITY OF MINNESOTA

Reasons for Rehospitalization May Be Require Interpretation to Be Attributed to Surgery

Rehospitalizations Within 30 Days of Discharge Following Mastectomy, by Length of Initial Stay and Reason for Rehospitalization



Other condition

- Reasons other than below

Possibly surgically related

- Nausea and vomiting
- Syncope
- Pneumonia
- Sepsis
- Urinary tract infection

Definitely surgically related

- wound infection
- seroma/hematoma
- pulmonary embolism or deep vein thrombosis
- other surgical complications

Warren JL et al. J Natl Cancer Inst. 1998

Complications of Chemotherapy

Medicare data are longitudinal, therefore useful for capturing chemotherapy complications—especially those requiring hospitalization

- Anemia (284.285.9)
- Bacteremia (790.7)
- Dehydration (276.5)
- Delirium (780.x)
- Diarrhea (787.91, 564.5)
- Fever (780.6)
- Infection (001.0-139.8)
- Sepsis (038.0-038.9)
- Stomatitis (528.0)
- Neutropenia (288.0)
- Thrombocytopenia (287.4)
- Unspecified adverse effect of systemic therapy (E93.31)

Measuring Chronic Complications

- May include
  - Pain
  - Fatigue
  - Nausea
  - Loss in function
- Most do not require hospitalization

Whether a Complication Appears on a Claim Varies by Condition

Self-Report of Chronic Complications vs. Conditions Found in the Medicare Claims for Men Treated with Radical Prostatectomy

		Medicare Claims vs PCOS (Gold Standard) %		
	Kappa	Sensitivity	Specificity	PPV
Complication				
Treatment of urethral strictures	0.76	83	95	77
Incontinence				
All codes	0.27	29	93	48
Procedures only	0.23	18	99	75
Big/moderate bother from incontinence	0.30	39	92	36
Impotence				
All codes	0.02	12	94	94
Procedures only	0.01	2	100	100

Potosky AL et al. Med Care. 2002

Late effects of treatment

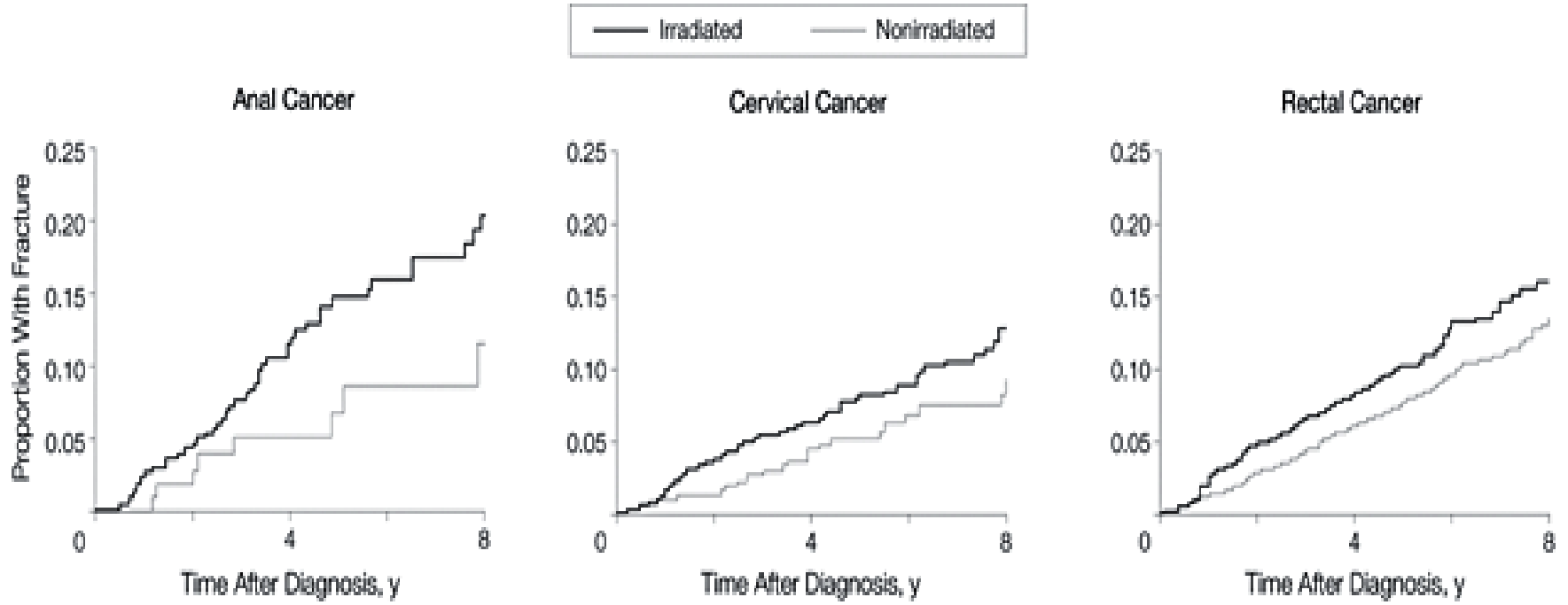
Longitudinal Nature of SEER-Medicare Data Make it Useful for Assessing Late Effects

- Need to be something significant (fracture, second cancer) that will be reported in claims or registry data
- Important to include comparison group

Example of Study of Late Effects of Treatment: Pelvic Fracture Following RT

- Elderly women are at increased risk of fractures
- Radiation can result in bone damage and may increase fracture risks
- Little is known about the risk of pelvic RT among elderly women
- SEER-Medicare data used to determine if women who undergo pelvic irradiation for anal, cervical, or rectal cancers had higher rates of pelvic fracture than women with pelvic malignancies who do not undergo irradiation

Time from Cancer Diagnosis to Pelvic Fracture, Up to 15 Years Later, by Treatment Among Elderly Medicare Women



No. at Risk									
Irradiated	399	182	75	1139	411	179	1317	505	176
Nonirradiated	157	61	27	466	242	102	2950	1470	587

Baxter, N. N. et al. JAMA 2005.

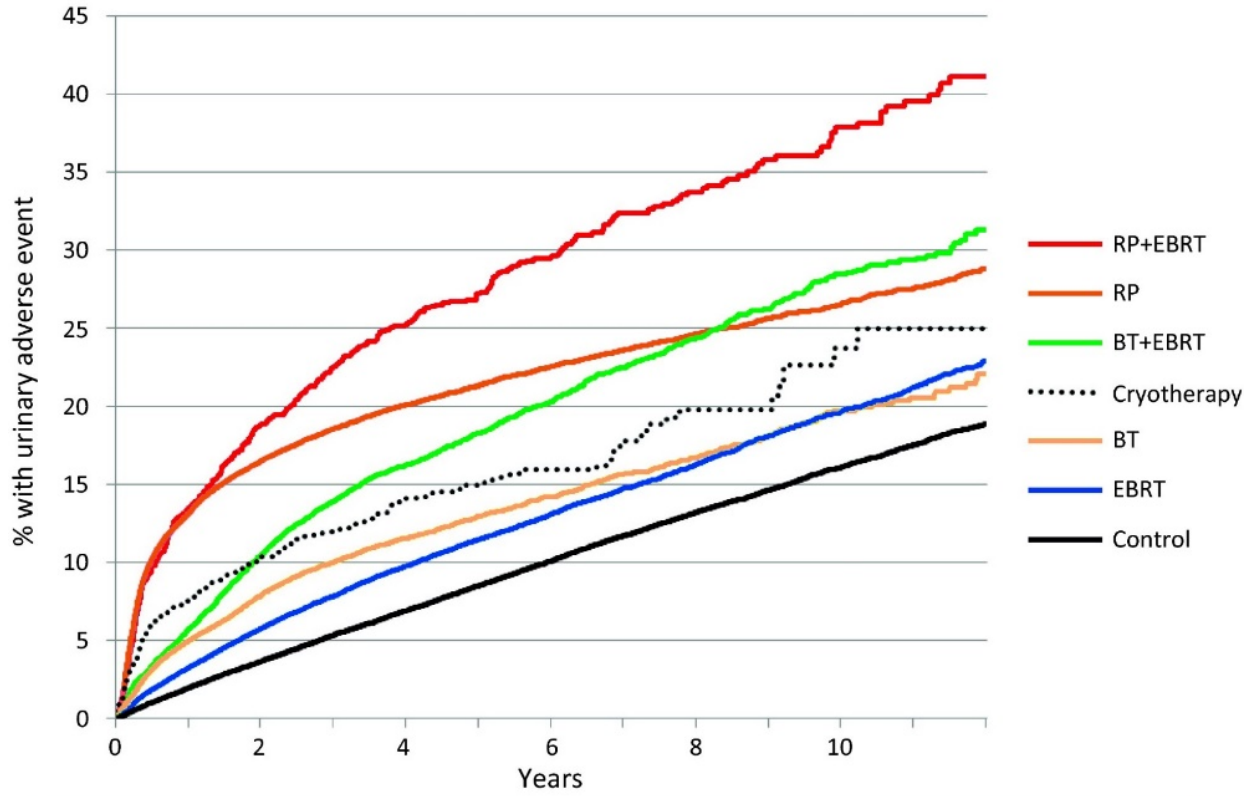
Example—long Term Risk of Urinary Adverse Events of Local Treatment for Prostate Cancer

- Carefully defined adverse events that could be ascertained using claims data
  - Common Terminology Criteria for Adverse Events (CTCAE)
    - From the National Cancer Institute
    - Grade 3 or 4 events
      - Required both a urinary adverse event (UAE) diagnosis code and correlating procedure code

CTCAE class (proportion of all UAEs)	Diagnosis	ICD-9 diagnosis code	Procedure names	CPT	ICD-9 procedure code
Bladder perforation (0.04%)	Bladder perforation	596.6	Cystorrhaphy	51860, 51865	57.81, 57.82
			Percutaneous abdominal drain	75989	54.91

Jarosek et al. European Urology 2015 67, 273-280

Claims Data Can Be Useful for Long Term Follow-up



	Number of individuals at risk					
	0 yr	2 yr	4 yr	6 yr	8 yr	10 yr
RP+EBRT	1557	897	607	399	251	163
RP	26 790	16 922	12 571	9040	6084	4030
BT+EBRT	11 835	7710	5134	3051	1275	456
Cryotherapy	2115	1096	537	213	72	–[a]
BT	44 318	27 438	18 049	11 006	5761	2974
EBRT	14 259	9529	6236	3380	1280	452
Control	144 816	77 348	48 457	31 287	20 729	13 262

[a]Cell masked for $n < 11$, in accordance with National Cancer Institute guidelines.

Conclusion Related to Using SEER-Medicare to Assess Adverse Events

- Studies that look at short term outcomes, especially those that require procedures or hospitalizations have a lot of potential
- Long term adverse events outcomes that are based on diagnoses only, especially only in the physician data, will be underestimated as outcomes
- Late effects of treatment will be in the data as long as the later events are significant.
  - Use of a comparison group is important

Comparative Effectiveness Research

Comparative Effectiveness Research (CER)

- Definition:
  - Systematic research comparing different interventions and strategies to prevent, diagnose, treat and monitor health conditions
- Purpose:
  - To inform patients, providers, and decision-makers about which interventions are most effective for which patients under specific circumstances

Why SEER-Medicare Data Might Be Considered for CER

- Elderly people are underrepresented in clinical trials
- SEER-Medicare has large numbers from multiple institutions
- Data are longitudinal

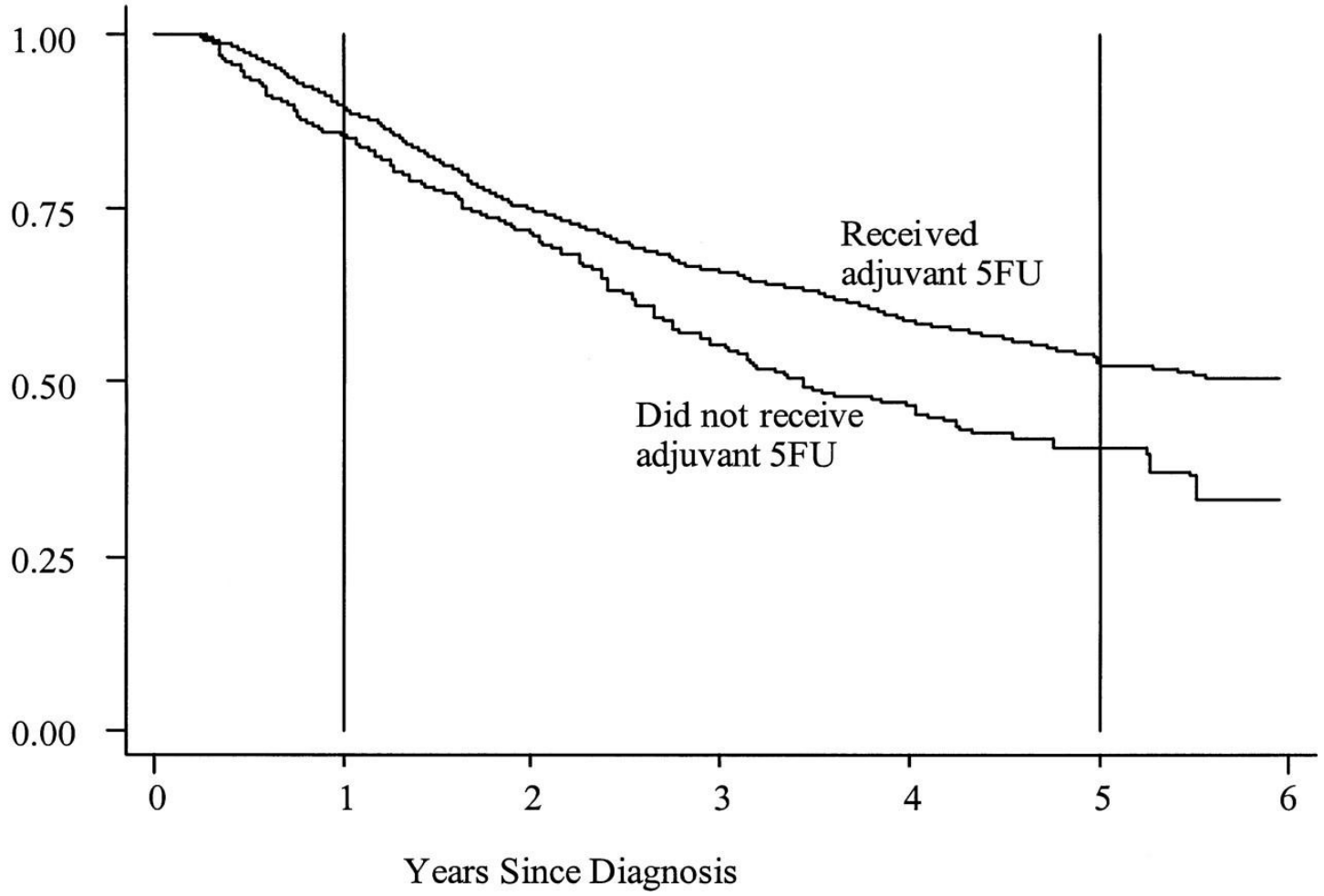
Why SEER-Medicare Data Might Not be Appropriate for CER

- People in SEER-Medicare are not randomly assigned to treatment
- Statistical adjustments may not be able to fix all aspects of non-random assignment
- Let's look at some examples

The Benefit of 5FU for Elderly Persons With Stage III Colon Cancer

- Adjuvant 5FU has long been the standard of care for persons with Stage III colon cancer
- Elderly persons are less likely to receive chemotherapy; no differences in toxicity noted in clinical trials
- Researchers used the SEER-Medicare data to compare outcomes between persons with Stage III cancer who did and did not receive 5FU

Kaplan-Meier Estimates for Overall Survival: Stage III Colon Cancer Age 65+ Treated With 5-FU and Propensity Score–matched Untreated Patients



Iwashyna, T. J. et al. J Clin Oncol 2002

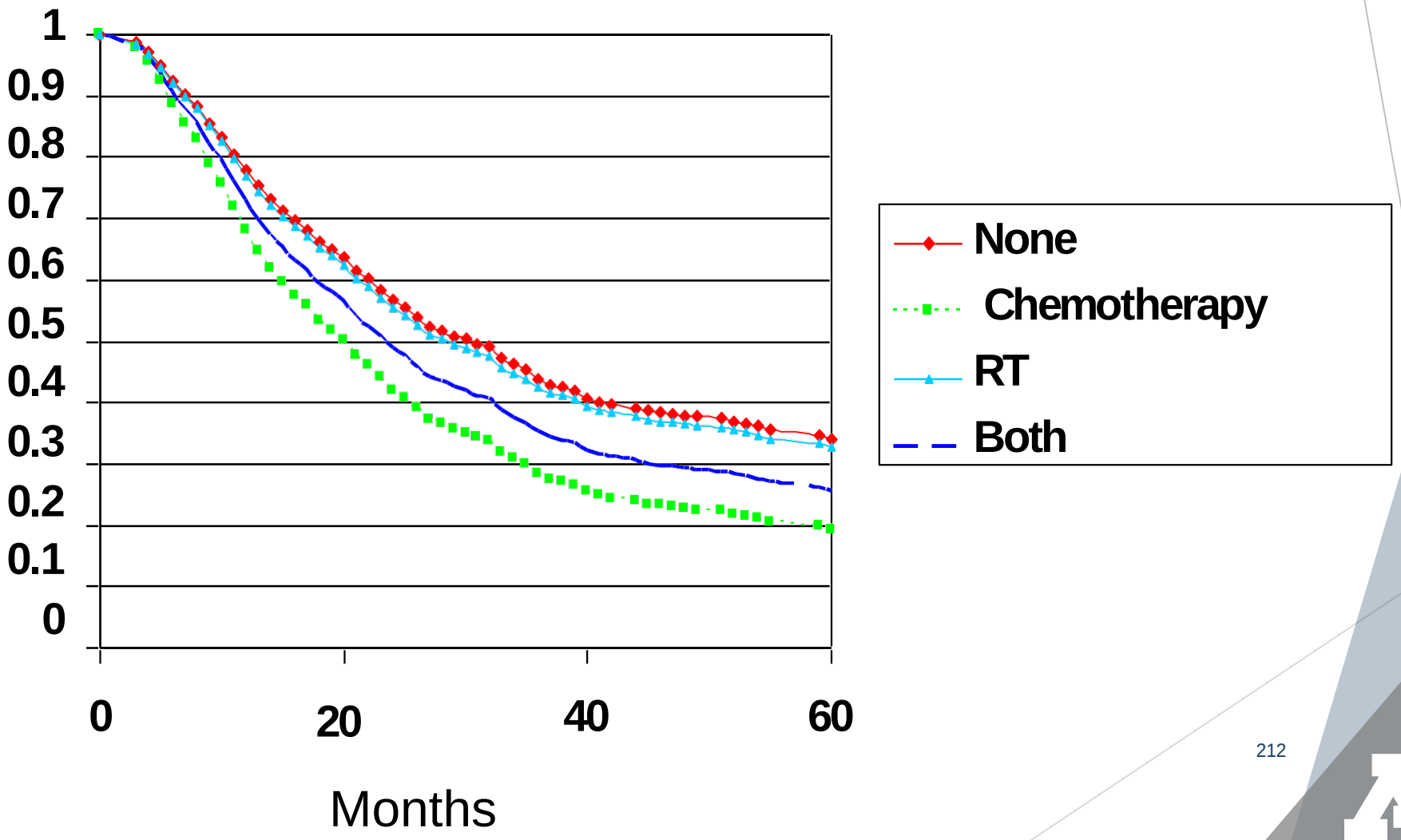
Kaplan-Meier Estimates for Overall Survival: Stage III Colon Cancer Age 65+ Treated With 5-FU and Propensity Score–Matched Untreated Patients

- Were comparison groups balanced?
- Or were treated patients healthier?
- If results are consistent with RCT, does this prove effectiveness in the "real world"?
- Would we believe results that vary from RCT (and would we change practice)?
- What if we have no trial data to compare with? Example of pancreatic cancer analysis

CE of Chemoradiation for Locally Advanced Resected Pancreatic Cancer

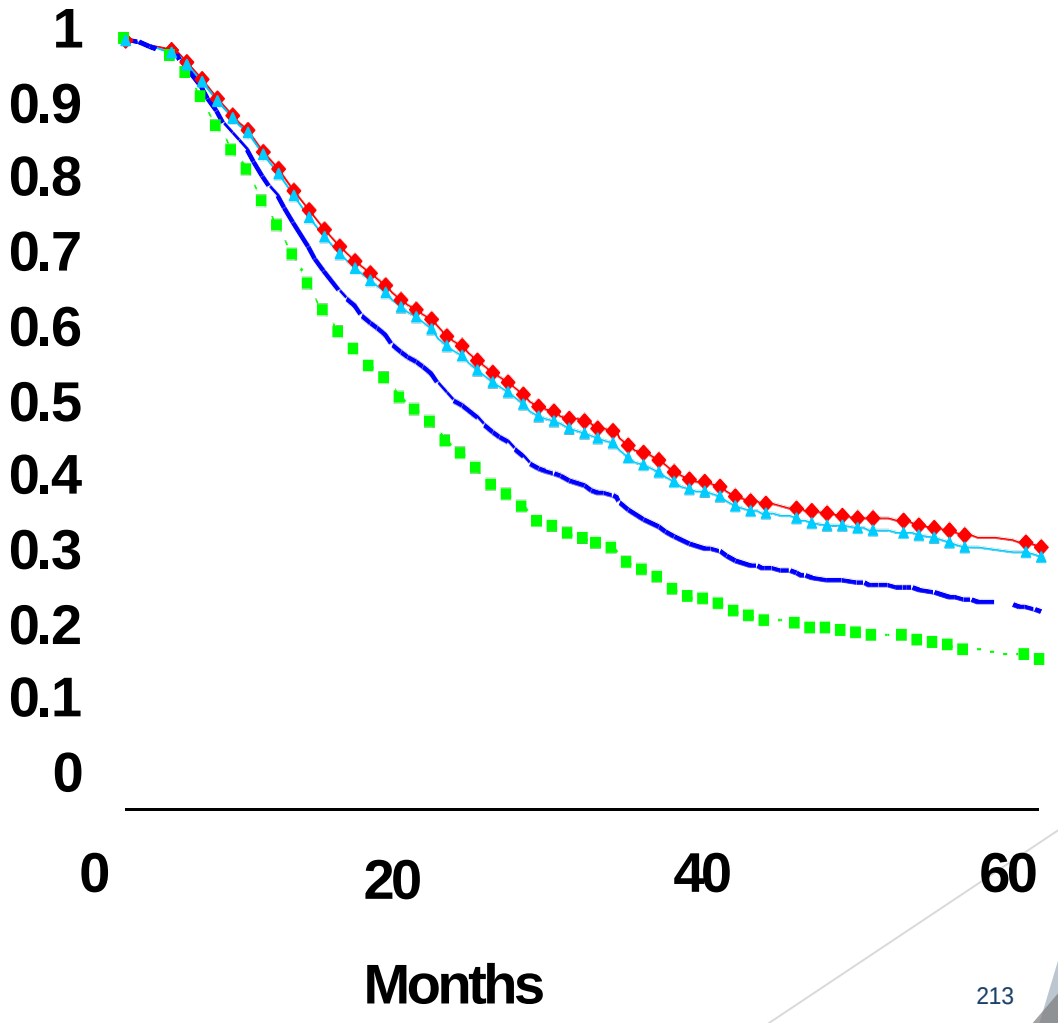
- Chemotherapy alone is the recommended standard of care for patient with locally advanced resectable pancreatic cancer
- An earlier clinical trial (GITSG) reported benefit of RT for locally advanced disease.
- Another trial (ESPAC-1) found RT is detrimental, leading to debate about the risk/benefit of RT
- NCI researchers used SEER-Medicare data to compare survival following different treatments among patients with local advanced, resected pancreatic cancer

Survival Among Persons With Node-negative Surgically Resected Pancreatic Cancer by Type of Treatment, Persons Age 65-69, White, Comorbidity Score=0



Using Observational Data to Assess Treatment Benefit is Complicated

- What do the data tell us about which treatment is most beneficial to patients with early stage pancreatic cancer?
- The data are counter-intuitive, with patients who received no treatment having the best outcomes
- These data suggest that there are unmeasured differences between treatment groups



Conclusions About Using SEER-Medicare Data to Assess CER

- Using SEER-Medicare data to assess the effectiveness of cancer treatment is perilous, results could be misleading
- Statistical methods (Instrumental Variable Analysis (IVA)/propensity scores) may remove biases compared to traditional methods.
  - However, these methods do not completely control for observed and unobserved differences between groups in analyses based on secondary data
- Confirmation/validation required via prospective trials when possible
- Use of these data is best for epidemiologic investigations and hypothesis generation, rarely for guiding clinical decisions

Recurrence

Measuring Recurrence/Disease Progression

- There is great interest in assessing the impact of cancer treatments beyond observed survival
  - Most patients present with local-regional disease
  - Patients are living longer with cancer
  - The cost of treating cancer has risen significantly
  - An important aspect of treatment success

Medicare Data *Could* Be Used to Infer Recurrence

- Additional treatment usually given for recurrence (more surgery, chemo, RT)
  - Treated recurrence
- ICD-9 or 10 diagnosis codes for metastasis
- Entry into hospice/palliative services
- Death from cancer

Treated Recurrence

- Some investigators have attempted to identify recurrence/progression through treatment
- Not all patients with recurrence/progression will opt for additional treatment, especially among elderly
- For SEER-Medicare breast and colorectal patients with Stages II or III disease, about 1/3 of patients received additional surgery, chemotherapy or RT when they first recur. Patients 75 and over were significantly less likely to receive additional treatment.
- Patients treated for recurrence with oral drugs will be missed

For Studies Assessing Treated Recurrence

- It is helpful to look for treatment gaps
  - Caveat: is chemotherapy given 4 months following surgery adjuvant?
- Hard to gauge progression for cancers that are treated with "watch and wait" approach such as prostate cancer
- Watch definitions -ex. recent manuscript only used RT/chemo six months after diagnosis as evidence of recurrence. The authors classified the people who did not have RT/chemo as having "disease free survival" when they really had "treatment free survival"

ICD-9 or 10 diagnosis Codes for Metastasis

Bone metastases at diagnosis Percent found in Medicare claims in the same month of diagnosis (of those identified by SEER)				
			Medicare Claims	
YEAR	Sample Size	**SEER**	**Absent**	**Present**
2010	58,953	Absent	98.2%	1.8%
	3,625	Present	62.4%	37.6%
2011	58,264	Absent	98.4%	1.6%
	3,706	Present	59.7%	40.3%
2012	56,673	Absent	98.6%	1.4%
	3,833	Present	61.0%	39.0%
2013	54,527	Absent	99.0%	1.0%
	3,800	Present	61.7%	38.3%

Do Metastasis Codes on Claims Perform Better at a Time After Diagnosis?

- Study #1 (Hassett et al. Medical Care 2012). Used CanCORS/Medicare and HMO/Cancer Research Network data to identify patients with claims for metastasis or chemotherapy after initial treatment. They concluded:
  - Metastasis codes and chemotherapy codes could not identify recurrent cancer without risk of misclassification.
  - No code-set was highly sensitive and highly specific
  - Findings based on existing algorithms should be
- Codes from claims should be interpreted with caution

Do Metastasis Codes on Claims Perform Better at a Time After Diagnosis?

- Study #2 (Nordstrom et al. Pharmacoepidemiology Drug Safety 2012). Used outpatient EHR data linked to medical and pharmacy claims. They compared metastasis identified from claims with what was reported on the EHR
- Findings-
  - PPV ranged from 0.75 to 0.86
  - Specificity 0.75- 0.97
  - Sensitivity 0.60-0.81
- They concluded:
  - "Results suggest that accurate ascertainment of metastatic status may require access to medical records or other confirmatory data sources."

Entry Into Hospice/Death from Cancer

- Hospice diagnosis codes
- Cancer Cause of Death
- If using multiple measures of recurrence, remember that treated recurrence will show up earlier than untreated recurrence
  - Will introduce bias, amount may depend on time from recurrence to death
  - Disparities in treated recurrence will increase bias

Conclusions About Using SEER-Medicare Data to Assess Recurrence

- Using SEER-Medicare data to assess recurrence is perilous, results could be misleading
  - Watch for bias introduced by recurrence measures
  - Be careful about terminology—'treated recurrence'
- Use of these data is best for epidemiologic investigations and hypothesis generation, rarely for guiding clinical decisions

Mortality

Outcome: Measuring Cancer Deaths/Survival

SEER	Medicare
Month and year of death	Date of death
SEER dates of death reported though year of case ascertainment	Medicare date of death reported through final year of claims

- High concordance between the two
  - But be careful of using date of death from Medicare and cause of death from SEER without first confirming concordance.
- Because there is no cause of death on the Medicare data, you will not have cause of death for non-cancer cases.
- If you want to attribute cause of death to cancer, limited to SEER

Cause of Death Reporting

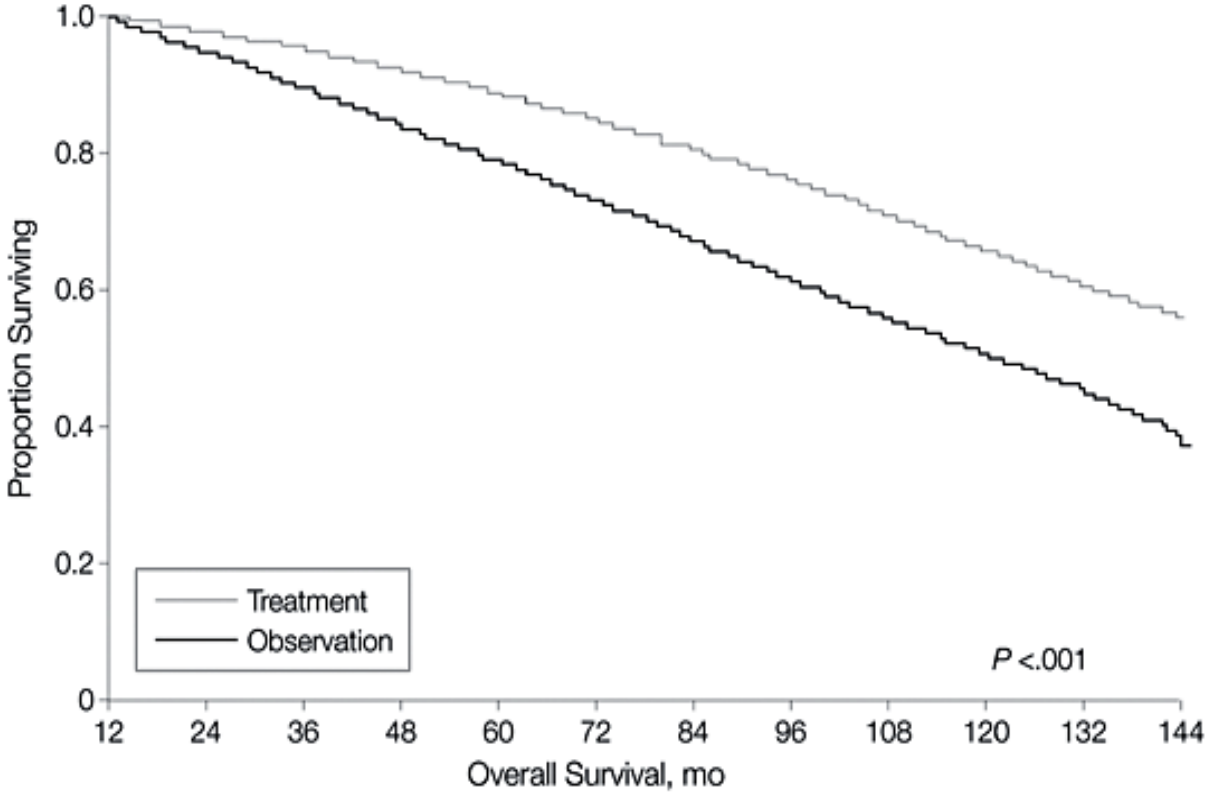
- Many elderly have more than one cancer making it difficult to attribute a cancer death to a specific cancer.
- Cause of death may be miscoded
  - Ex: a patient with lung cancer and brain metastasis may be miscoded as brain cancer being cause of death
    - Suggested approach--Take patients with only one cancer and assume that they died from that cancer if any type of cancer is reported as COD.
  - Worse: what do we do with someone with lung cancer who died of pneumonia?

Why the Correct Measure of Death Matters--An Example

- There is ongoing uncertainty regarding the best treatment for localized prostate cancer in elderly men
- Most men with localized prostate cancer will not die from their disease
  - Researchers used the SEER-Medicare data to assess 5- and 10-year overall survival comparing men who were treated for their prostate cancer (RT or RP) with those with expectant management. They used propensity scores to account for differences between the two groups

Importance of Differentiating Between Dying With Vs Dying of a Cancer

Adjusted Overall Survival Curve by Treatment for Men Age 65-80 Diagnosed With Prostate Cancer: 1991-1999



No. at Risk												
Treatment	32022	31378	30546	26519	22630	18724	15148	11674	8755	5805	2522	226
Observation	12608	11986	11308	9696	8124	6587	5163	3769	2608	1527	685	82

However, There is Another Side to the Story

Group	N	10-year Overall Deaths	10-year Prostate Cancer Deaths
Treatment	7639	23.8%	1.9%
Observation	4663	37.0%	2.5%

Conclusions About Using SEER-Medicare Data to Assess survival

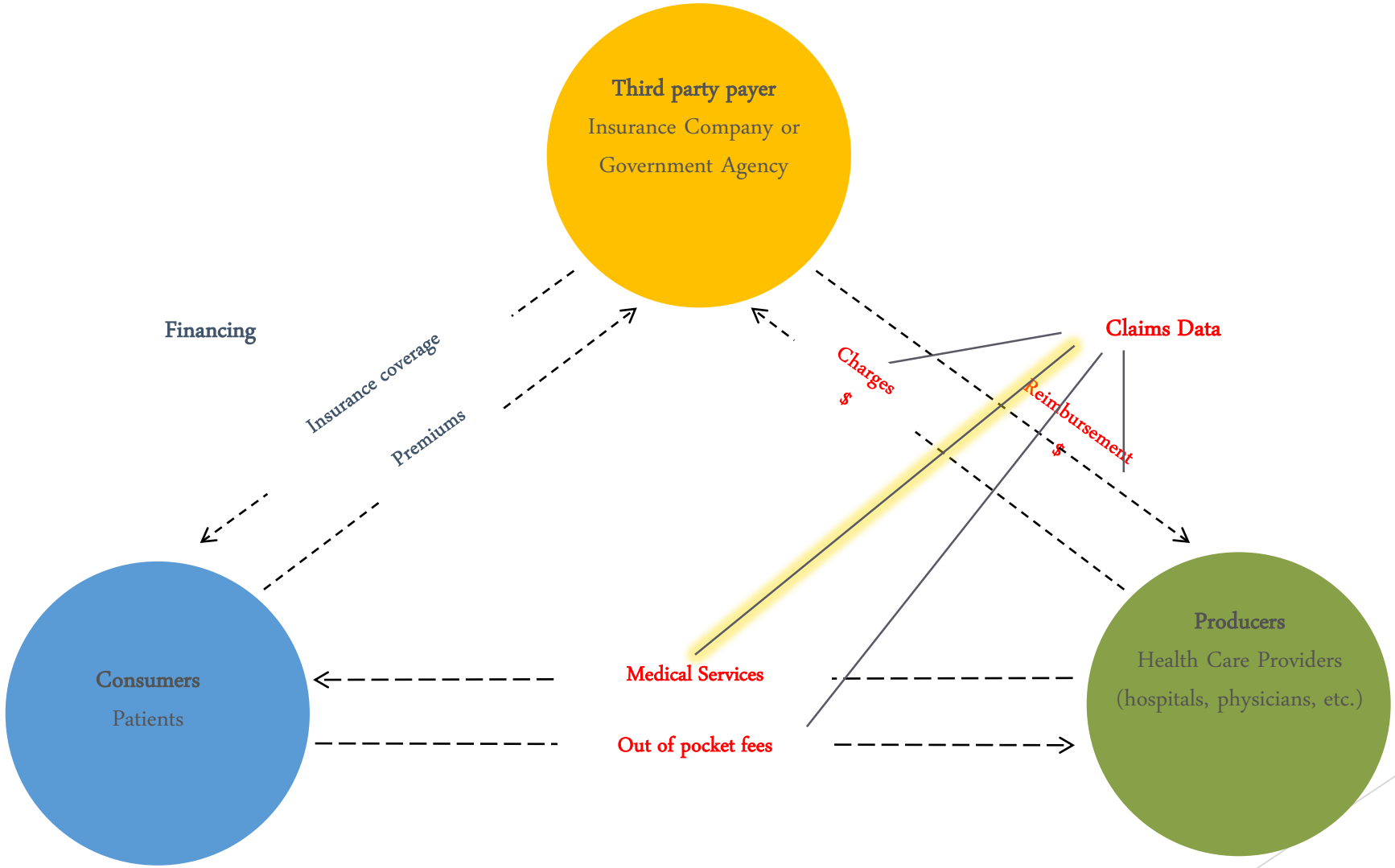
- SEER registries and the Medicare program both contribute information about death.
  - Which you choose will depend on your goals
- If using the two in combination, remember the limitations of each
  - For example, remember that cause of death information ends two years before Medicare death information ends

Costs and payments

Many Sources of Information About Money

- Charges
- Payments
  - By Medicare
  - Leftover after Medicare pays (patient responsibility)
- "Best" Type of Cost Estimate and Method Depends on Underlying Research or Policy Question

Healthcare Payment Overview

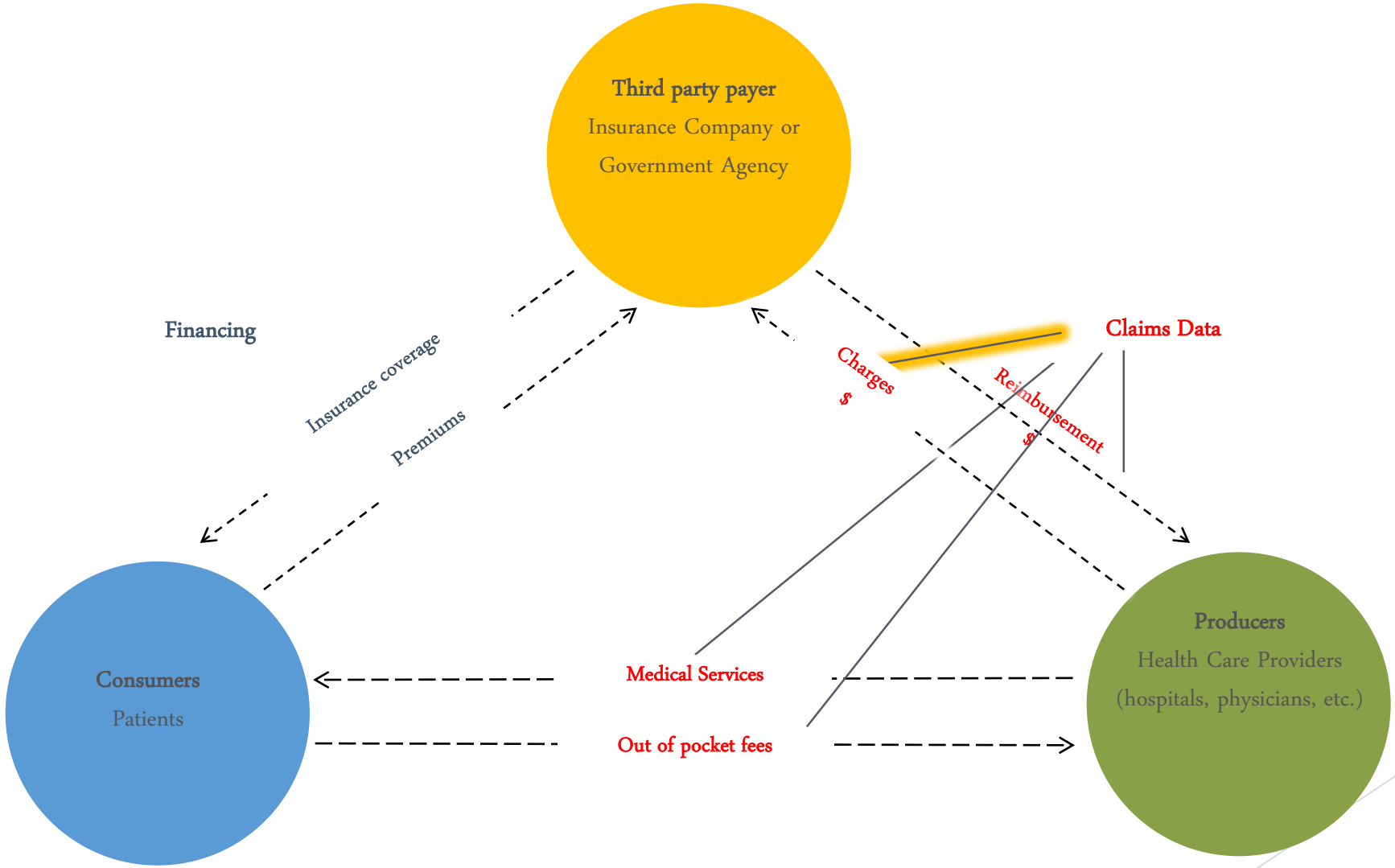


SCHOOL OF PUBLIC HEALTH
UNIVERSITY OF MINNESOTA

Basis of Payment Varies by Service Type

	ICD-9/10 diagnosis	ICD-9/10 Procedure	CPT/ HCPCS	Revenue Centers
MedPAR	*	*		
Physician/NCH(Carrier) DME			*	
Hospice			* (for MD services only)	*
Hospital Outpatient			*	*

Healthcare Payment Overview



Charges

- Charges: Set by provider
  - For facilities, usually a large discrepancy between payment and charge
  - For physicians, Medicare determines what the provider is allowed to charge for a service (aka ‘Allowed charge’). Payment of the allowed charges come from two sources:
    - Medicare
    - Beneficiary: Co-payments, coinsurance, and deductibles, (may be paid by coinsurance such as Medigap)

Charges from Provider

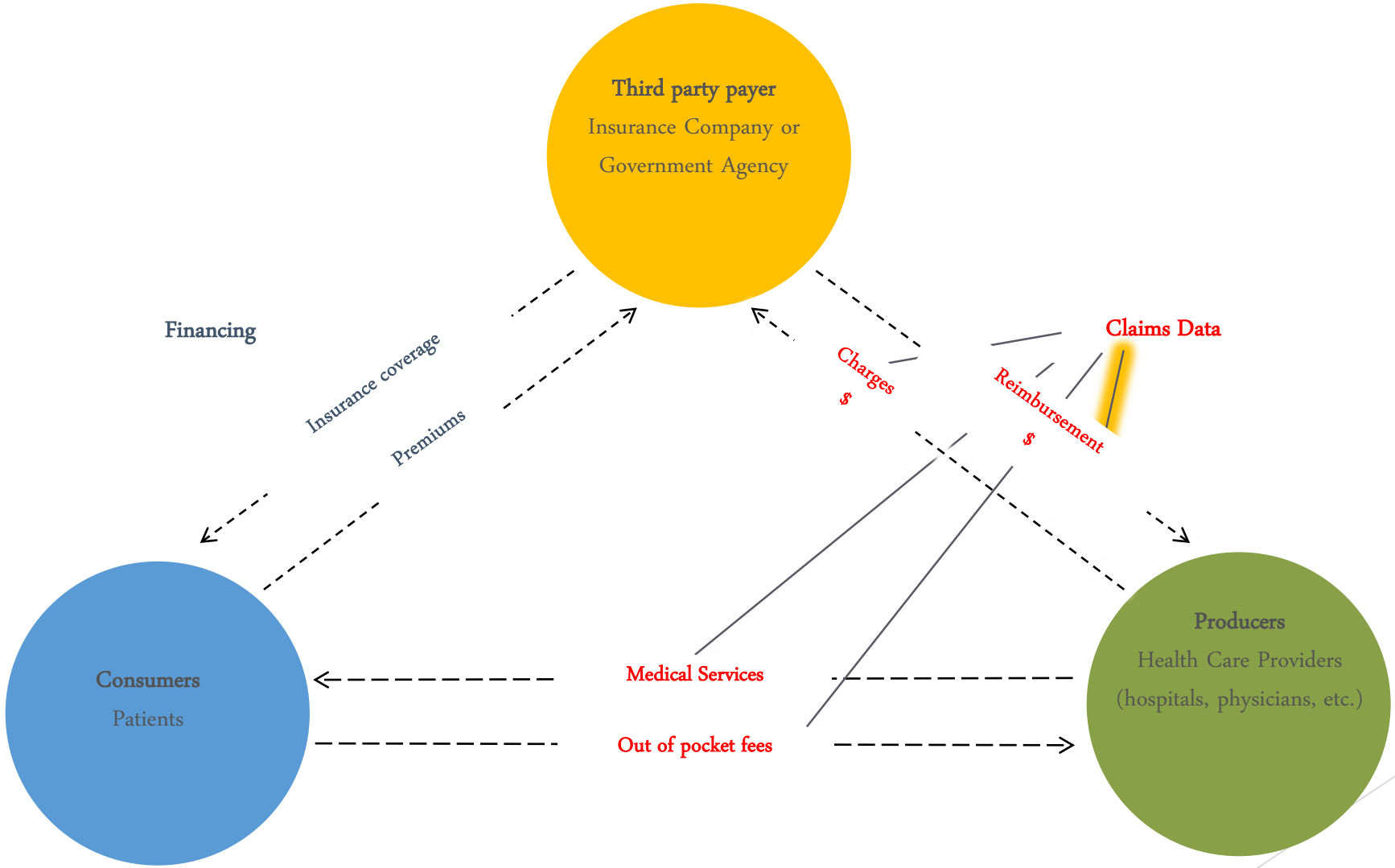
- Observable on claims, but should not be used without adjustment
  - MedPAR: Total charge amount

Note: Includes discharges from Hospitals located within the 50 United States and District of Columbia

DRG Definition	Total Discharges	Average Covered Charges	Average Total Payments	Average Medicare Payments
191 - CHRONIC OBSTRUCTIVE PULMONARY DISEASE W CC	115,897	$27,016.90	$6,947.29	$5,672.03
192 - CHRONIC OBSTRUCTIVE PULMONARY DISEASE W/O CC/MCC	70,430	$21,142.92	$5,501.10	$4,200.94
193 - SIMPLE PNEUMONIA & PLEURISY W MCC	135,974	$40,481.27	$10,222.00	$8,810.66
194 - SIMPLE PNEUMONIA & PLEURISY W CC	150,770	$28,084.04	$7,252.70	$5,875.41
195 - SIMPLE PNEUMONIA & PLEURISY W/O CC/MCC	55,557	$20,721.30	$5,353.19	$4,008.63

https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/Inpatient.html

Healthcare Payment Overview



SCHOOL OF PUBLIC HEALTH
UNIVERSITY OF MINNESOTA

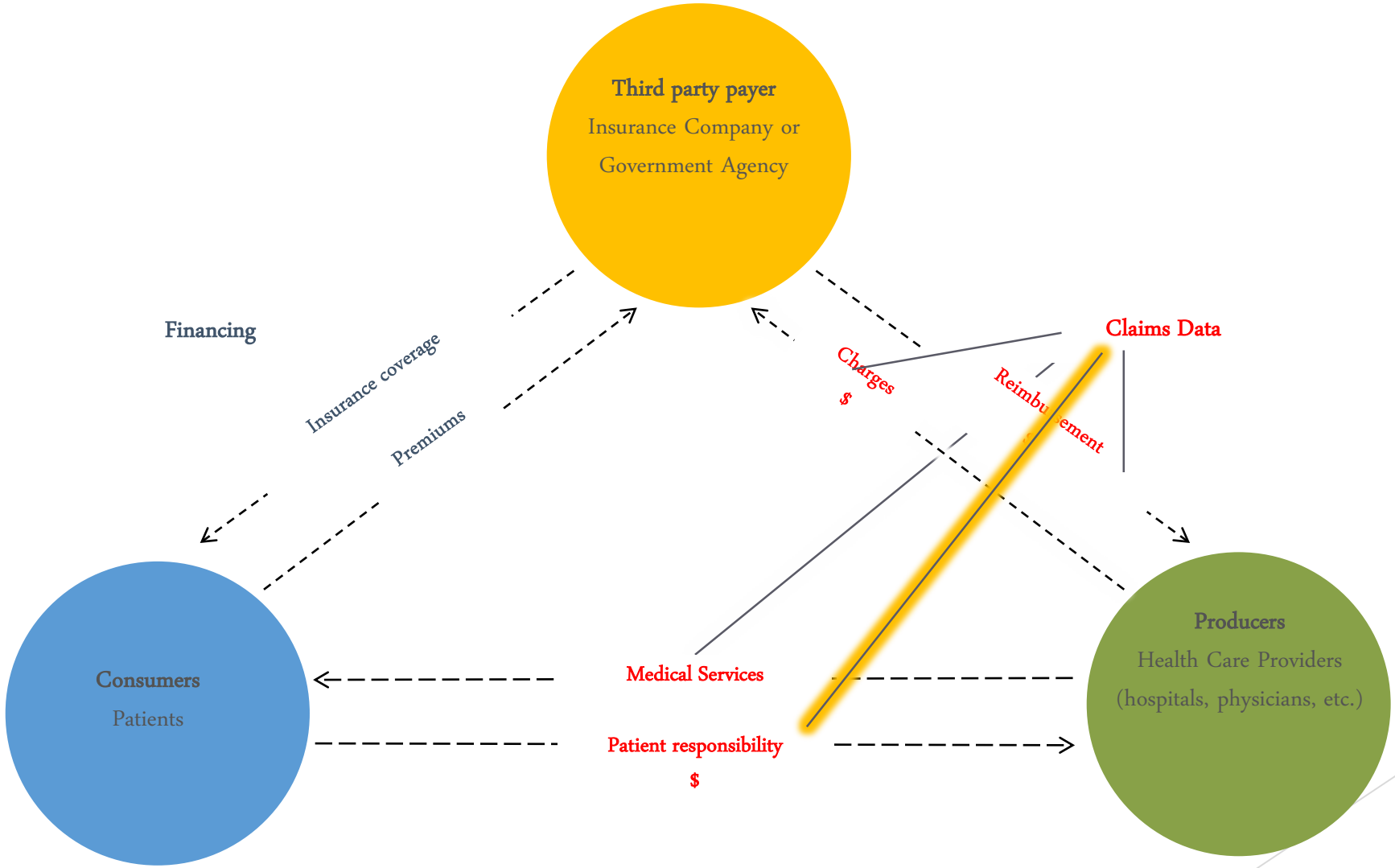
Reimbursement – from Medicare

HOSPITALS

- Acute Inpatient Perspective Payment System (IPPS)
  - Paid based on MS-DRG

- Operating base payment rate (national) weighted based upon MS-DRG
  - Adjusted for geographic factors
  - Adjusted for case mix
  - Policy adjustments
    - IME (indirect medical education)
    - DSH (disproportionate share payment)
  - Adjustment for transfers
    - Full LOS
    - Short LOS and transfer

Healthcare Payment Overview



Beneficiary Responsibility

- Beneficiary co-payment and deductible fields contain the amount the beneficiary owes.
  - It does not contain the amount the beneficiary actually paid.
- There is no way to determine how much of the beneficiary responsibility was actually paid to the provider.
- There is also no way to determine whether the beneficiary had other coverage (e.g. Medigap, Medicaid) that paid part or all of the beneficiary's responsibility

Patient's Share of the Payment

Part A (Hospitalization Insurance)

- Deductible
  - $1,340 for each benefit period

- Coinsurance
  - 1 through 60 days: $0 coinsurance
  - 61 through 90 days: $335 per day coinsurance
  - 91 days and beyond: $670 per day coinsurance for each lifetime reserve day (You get 60 lifetime reserve days)
  - After lifetime reserve days are used up: You pay all costs
  - SNF coinsurance (days 21-100): $167.50 per day

Patient's Share of the Payment

Part B (Supplementary Medical Insurance)

- Deductible (annual)
  - 2018: $183
- Copayment
  - Medicare generally pays 80%
  - 20% is the patient's responsibility
  - NCH/Carrier File: Allowed Charge Amount
    - Generally, but not always, equal to copayment + Medicare payment
  - Outpatient file
    - Copayment may be set by the Hospital Outpatient Prospective Payment (HOPPS) APC (will not always be 20%)

Costs and payments

Challenges with Estimating Cancer Related Costs

You Will Not Find One Bill for All the Costs Associated With Treatment

- Multiple components of cost
  - Cancer-directed surgery
  - Chemotherapy
  - Radiation oncology
  - Other hospitalizations
- There is no one-to-one relationship between these components and Medicare claims
  - Many services are billed in a series of components
    - Surgeon, anesthesiologist, pathologist, facility (MedPAR or outpatient) will often bill separately

Separating Services and Associated Costs Related to Cancer from Other Healthcare

- Identify specific procedures and services that are clearly cancer specific – e.g., surgical removal of a tumor, chemotherapy, etc.
- Case-control approach, compare costs for cancer cases to matched controls without a cancer diagnosis
- Identify period where majority of care is cancer- related (e.g., cancer directed surgery)
- Statistical models

Payment Rules Vary by Service Type

- Hospitals have per-spell of illness deductibles and co-payments
- Outpatient services have an annual deductible as well as a co-payment
  - Some services have no co-payment, e.g.:
    - Cancer screening
    - Immunizations
    - Bone mass measurement
    - Yearly ‘wellness’ visit (new)

MedPAR--Inpatient Stay

- Total payments from Medicare are calculated as reimbursement amount PLUS total pass through amount
- Primary payer amount is the amount paid by another insurance (before Medicare)
- Beneficiary responsibility:
  - Sum of co-payments and deductibles
- Total payments to hospital: need to add together Medicare payments, primary payer amount and copayment/deductible

Physician (NCH/Carrier) Payments

- Payments in the NCH/Carrier file are based on service. For each service, or line item, there is an allowed amount which is divided into the Medicare and patient responsibility
- Line item payment by Medicare:
  - Line Payment Amount + Line Interest Amount
- Earlier linkages: Line Payment Amount
- Line item beneficiary responsibility:
  - Line Beneficiary Part B Deductible Amount + Line Coinsurance Amount

NCH/Carrier Payments

- Total claim payment by Medicare:
  - Claim Payment Amount + Sum of line interest amounts
- Earlier linkages: Claim Payment Amount
- Total beneficiary responsibility for claim:
  - Sum of line item co-payments and deductibles
- Total payment due provider:
  - Medicare payment + Beneficiary co-pay/deduct + Claim Primary Payer Amount
    - This will usually equal the “Allowed Amount”
- Note: Claim Payment Amount repeats for each line item on the claim. Be careful to count it once per claim.

Payments on the Outpatient File

- Payment in the Outpatient file are also based on service, referred to as ‘Revenue Centers’. This is akin to Line Items.
- Each line in the Outpatient file is a ‘Revenue Center’ within a claim.
  - Each claim contains at least two records, one of which is a summary record.
    - Summary record contains Revenue Center = 0001

Payments on the Outpatient File- Part II

- For services on or after July, 2000, the Outpatient file contains Revenue Center payments. Prior to that, the file contained total claim payments only.
- Revenue Center payment by Medicare:
  - Revenue Center Payment Amount*
- Revenue Center beneficiary responsibility:
  - Revenue Center Patient Responsibility Payment Amount
- Earlier linkage years: Not available

*Revenue Center = 0001 record will contain $0.00

Payments on the Outpatient File- Part III

- Total Claim Payment by Medicare:
  - Claim Payment Amount
  - **This repeats on every line; be careful not to double-count**
    - Earlier linkages: Claim Payment Amount, but only when Record Count field='0001'
      - BEWARE! If your analysis includes only select line items (e.g. chemotherapy), the record containing Claim Payment Amount may be absent.
- Total Beneficiary responsibility:
  - Beneficiary Part B Deductible Amount + Beneficiary Part B Coinsurance Amount + NCH Beneficiary Blood Deductible Amount
- Total Payment due to Provider:
  - Medicare Payment + Beneficiary Responsibility + Primary Payer Claim Paid Amount

Aggregate Cost Estimates

- Aggregate cross-sectional estimates in specific year useful for policy and program planning
  - Current burden (e.g., $103.6 billion in U.S. in 2006)
  - Proportion of program costs (e.g., x% of Medicare payments)
  - Future trends in incidence, survival, and costs
- Evaluate specific services or components of care in a specific year
  - Hospitalizations
  - Chemotherapy
- Evaluate trajectory in spending of a particular component of care
  - End-of-life

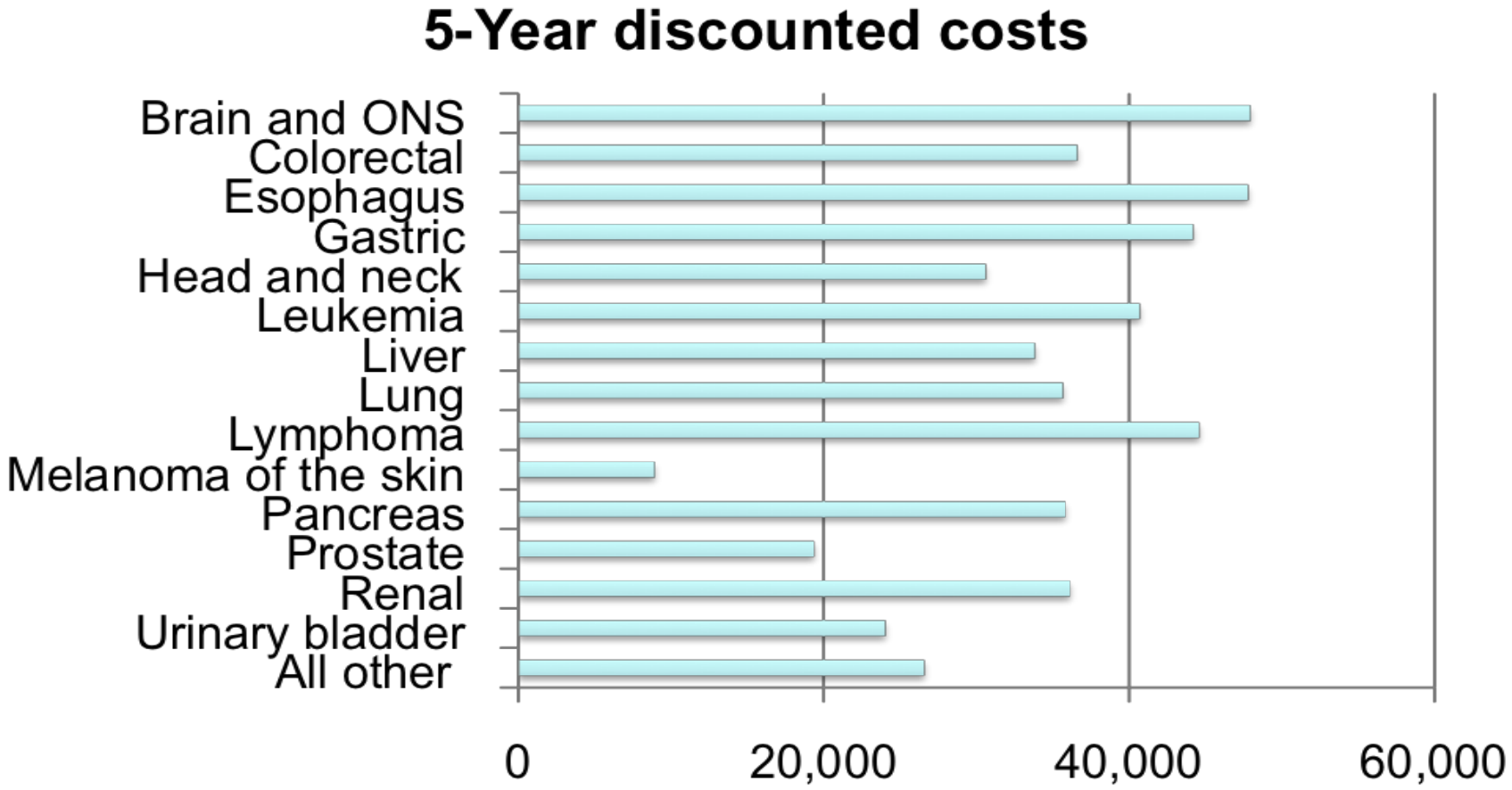
Estimating Components of Cost in First Year Following Diagnosis

- Components of cost
  - Cancer-directed surgery
  - Chemotherapy
  - Radiation oncology
  - Other hospitalizations
- There is no one-to-one relationship between these components and Medicare claims
  - Many services are billed in a series of components
- Hierarchical approach

Longitudinal Cost Estimates

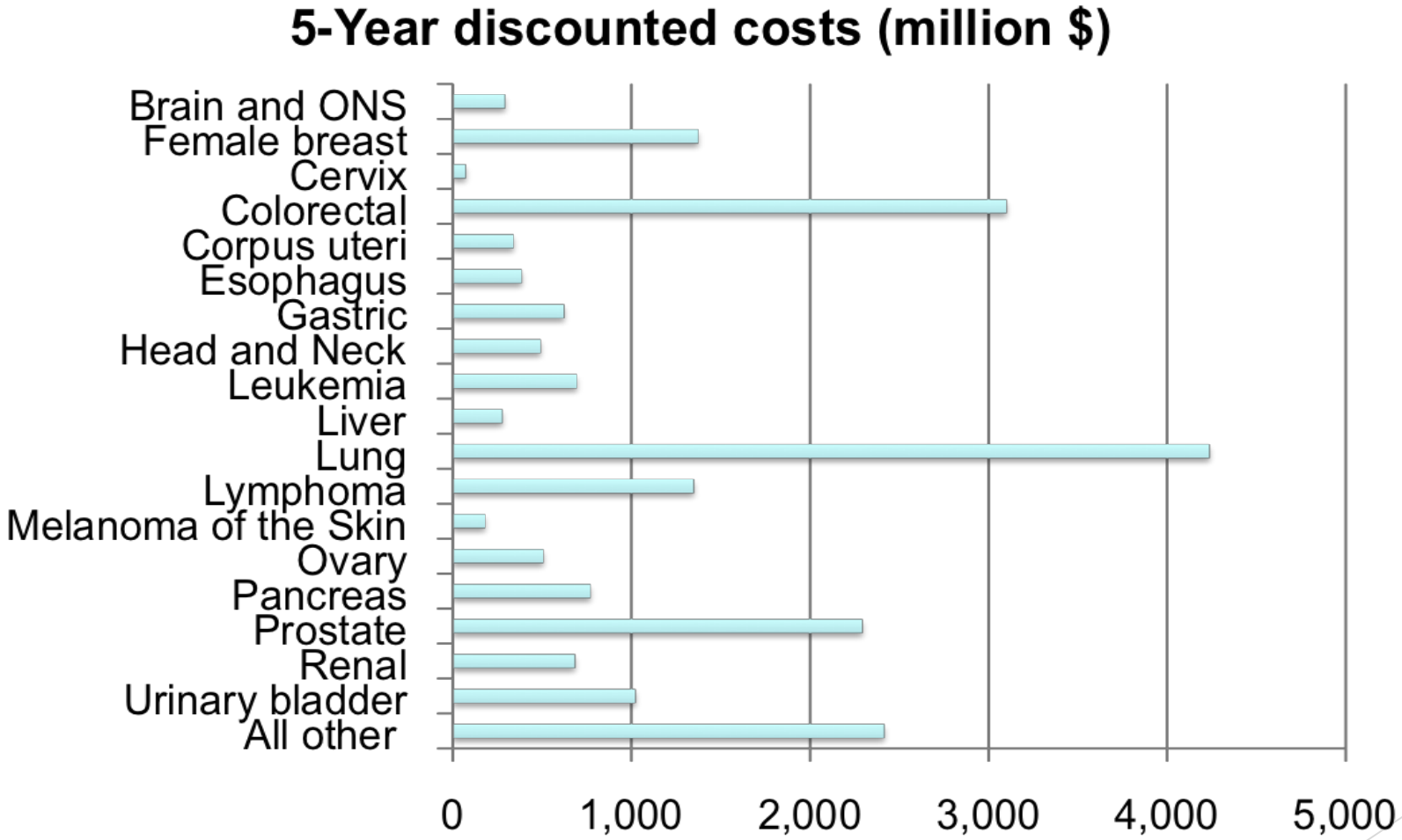
- Longitudinal per-person estimates useful for cost- effectiveness analyses of cancer prevention, early detection, treatment
- Can also be used to assess
  - Specific services or components of care
  - Care trajectory
- Stratified estimates
  - Stage of disease at diagnosis
  - Treatment-specific (e.g., type of breast surgery)
  - Provider-type (e.g., type of surgeon for ovarian cancer)
- Can be aggregated for newly diagnosed in a specific year

Mean 5-Year Net Costs of Care in Elderly Male Cancer Patients (in 2004 $)



SOURCE: Yabroff et al., J Natl Cancer Inst 2008 ;100(9):630-641.

Aggregate 5-Year Costs of Care in Elderly Medicare Cancer Patients Diagnosed in 2004 (in 2004 $)



SOURCE: Yabroff et al., J Natl Cancer Inst 2008;100(9):630-641.

Cost effectiveness Analysis—Measures of Effectiveness

- **Quality adjusted life years (QALYs) -**
  - Not in Medicare claims data directly
    - Mortality
    - Examine institutional claims and nursing home data to determine community residence
    - **Morbidity measures**
      - Length of stay, complications, readmissions for hospital or post-acute stays (CMS data: MedPAR, claims)

Measures of Effectiveness (continued)

- Comorbidity measures
  - CCW chronic conditions
  - Risk adjustment scores: use diagnoses (ICD) codes in claims files to calculate risk adjustment scores (e.g. Charlson, ACGs)

- Health & functional status measures are not in CMS claims data
  - E.g., # Activities of Daily Living (ADLs) or # Independent Activities of Daily Living (IADLs)

Measures of Effectiveness

- Health & functional status measures are not in CMS claims data
  - E.g., # Activities of Daily Living (ADLs) or # Independent Activities of Daily Living (IADLs)

- Assessment datasets: functional status during post-acute care stays
  - Minimum Data Set (MDS) – clinical assessment data for nursing home residents
  - Outcome and Assessment Information Set (OASIS) – assessment data for home care patients

Summary

- Unique opportunity to follow a population longitudinally for a variety of outcomes
- A lot of what we are interested in as researchers are captured in the administrative data processing
  - Death information
  - Payments
- But administrative data are not collected for research per se
  - Be thoughtful about what is being measured and how you are defining your measures

Questions?

Getting the Data & Publishing

Segment 6

SEER-Medicare Training 2019

Helen Parsons, PhD, MPH

Before Requesting SEER-Medicare Consider:

- The data must be used to answer a research question
- Each request for SEER-Medicare data is project-specific
- You may only use the files to work on a project as described in a proposal
- No Investigator can request the entire SEER-Medicare dataset
- SEER-Medicare data are not released outside of the USA
- Data are not immediately available

Approval Timeline

- 3-5 weeks for regular research requests
- 8-10 weeks for proposals involving restricted variables
- Data processing requires an additional 3-5 weeks after proposal approval

Each SEER-Medicare Request Requires a Minimum of Three Documents

- Application Form
- Signed Data Use Agreement
- Institutional Review Board Approval
- Request for Restricted Variables (if required)

Helpful Tips for Successfully Completing the Application Form

SCHOOL OF PUBLIC HEALTH
UNIVERSITY OF MINNESOTA

Section 1: Contact Information

- Principal Investigator (PI) is the person responsible for the appropriate acquisition, use, publishing and destruction of the data upon project completion
- Students and Fellows are not allowed to be PI
- Assistants/Fellows/Students can be listed as an alternate contact to receive project-specific correspondence

Section 2a/b: Project Title & Overview

- Brief (2 sentence) overview of the proposed project
- Must describe a **RESEARCH** question to be studied
  - Why?
    - HIPPAA allows for the healthcare data to be released for research purposes so project must describe a research question to be answered
    - Must create generalizable Knowledge

- Examples

Research

"To study the association between X and Y"

"To understand the role of X on Y"

Not Research

"Show how use of procedure X varies by geography"

"Develop an algorithm to classify complications"

"To develop a tool to …"

More Examples HERE

Section 2c: Cancer Sites Being Requested

- Consider the Minimum data Necessary Rule when Selecting Cancer Sites
  - "Limits the use or disclosure of, and requests for, protected health information to the minimum necessary to accomplish the intended purpose"
- Request only those cancer sites necessary to answer a research question
- Researchers are not allowed to request data all cancer sites for a proposal
  - Alternatively, consider requesting:
    - Top 3-5 most common cancers
    - 5% All Cancer Diagnosis File

Section 2d: Description of the Project (I)

- No more than 5 pages total
- Consider how you might present a brief proposal for research funding
- Need to provide a brief background/motivation for the project
  - Examples: Studies have shown that….; However, no research has examined …
- Re-state each of the research questions to be answered
- Briefly describe each of the key inclusion/exclusion criteria for your study
  - Common Criteria for those “likely to have complete claims” include:
    - Cancer site/Stage
    - Age at diagnosis (e.g., 66+)
    - Part A/B Enrollment
    - HMO Enrollment
    - Diagnosed at autopsy/on death certificate

Section 2d: Description of the Project (II)

- Need to describe all planned covariates/outcomes to be used in analyses and how you will define them
  - Codes (e.g., CPTs, ICD) that will be used
  - Include definition of the covariates/outcomes (e.g., Chemotherapy will be defined as presence of a claim for CPT codes XX-XX within X months of diagnosis using files X,Y, Z)
- Consider tables for summarizing measures to be used
- For helpful resources used to identify measures in SEER-Medicare:
  - SEER-Medicare Website
  - PubMed
  - Medical Care Special Issue

Section 2d: Description of the Project (III)

Remember, the following are not available or reliably captured in the analyses:

- Cancer Recurrence
- Physiology lacking:
  - BMI, Blood Pressure, Pulse, etc.
- Test results not included:
  - Lipid panel, Angiography, Pathology, etc.
- Behavioral information under-reported/not available:
  - Smoking, Alcohol use, Exercise, etc.

- For More Information, see Measures that are Limited or Not Available in the SEER-Medicare Data

Section 2d: Description of the Project (IV)

- In addition to descriptions of cohort development, covariates and outcomes, researchers must describe how EVERY SEER-Medicare file requested will be used:
  - Examples:
    - The PEDSF file will be used to identify patient demographics, identify inclusion/exclusion criteria including….
    - The MedPAR file will be used to determine cancer surgery, re-hospitalizations…..
- Include short description of planned statistical analyses
  - Examples:
    - We will use chi-square analyses and t-tests to examine..
    - Logistic regression will be used…

Section 2d: Description of the Project (V)

- Personnel Involved:
  - Should include all individuals who will have access to individual-level data (i.e., students, project analysts)
  - Institutional Affiliation for each individual
  - Do not include personnel to be named

- Timeline
  - High level overview of planned study activities (data acquisition, cleaning, analysis, publication)
  - Per data use requirements, project timeline can not exceed 5 years (extensions may be granted with prior approval)
  - Remember, these data are complex; be realistic

Section 2e: Data Storage and Protection

Must Include

- **Specific location** of the data and where/how the data will be stored
  - Examples: Data will be stored on a password protected, encrypted department server, which is located….
  - NO CLOUD STORAGE IS ALLOWED
- Details on how the data will be protected from unauthorized access
- Storage/protection of the media you receive containing the original files
  - Examples: Locked cabinet in PIs office
- Assurances that no attempt will be made to identify individual patients, hospitals or physicians
- Assurances that publications and presentations of the data will not allow identification of patients, hospitals or physicians.

Section 2f: Funding Source

- REQUIRED for all applications
- If the funding source is a for-profit company (e.g., consulting firm, pharmaceutical company) a funding letter is also required
- Funding letter must state:
  - PI is free to work and publish findings without limitations by the funder
  - Must come from a person in authority on company letterhead

Section 2g: Restricted Variables

- Patient, provider, hospital and geographic (e.g., zip code) identifiers are encrypted in the SEER-Medicare data
- Most researchers do not need access to true identifiers to complete analyses as they are encrypted the same way across all files
- Note that the census tract and zip code files already include many geographic characteristics for an individual without the need to link to census data
- With additional approval, researchers can request actual:
  - Census tract of the patient
  - Zip code of the patient, physician or hospital
  - Unencrypted hospital provider numbers
  - SEER-specific variables (e.g., Oncotype DX)

  However,
  - Unencrypted physician numbers are NOT available for request
- You must state ALL files you intend to link with the SEER-Medicare data. Linking without approval is a violation of the DUA.
- Access to restricted variables requires approval by all SEER PIs and additional approval time

Section 3: Data Files Requested

- List all data files and years requested for the proposed study
- Double-check that the files listed match your proposed cohort description, covariates and outcome measures
- Consider the need to obtain pre-diagnosis information about individuals
  - Example: A researcher requests breast cancer diagnoses from 2005-2010. Need to request claims from 2004-2011 to identify comorbidities in the year prior to diagnosis and care/comorbidities at least one-year post-diagnosis.

Data Use Agreement

- Lists the terms investigators agree to in order to access and use data
- READ THE ENTIRE DOCUMENT
- Key Points Researchers Agree to:
  - Not using data for purpose outside of proposed research
  - No data sharing
  - Notify NCI in the event of a PI move
  - Appropriately securing data (e.g., no cloud storage)
  - Submit all manuscripts/publications prior to submission

Institutional Review Board Approval

- Study title and PI listed on the IRB document must match the study proposal submitted
- Many IRBs, including NIH's Office of Human Subjects Research, have determined that the SEER-Medicare data are exempt (CFR 46.104(4))
- For more information on describing the dataset, consult resources on the SEER-Medicare website:
  - IRB Approval & HIPAA Regulations
  - About the SEER-Medicare Database

Top Five Reasons a SEER-Medicare Application is not Approved

- The central purpose of the study is not cancer research.
- The proposed research involves data that may compromise the privacy or confidentiality of patients, providers, or institutions.
- The research question is not sufficiently detailed to determine if the proposed analysis is feasible.
- The SEER-Medicare data are not of sufficient quality or completeness to provide accurate data to address a specific research question or aim.
- Missing application components (e.g., unclear description of key variables, incomplete discussion of how files will be used, unclear data storage descriptions)

Helpful Tips for Publishing Using SEER-Medicare Data

SCHOOL OF PUBLIC HEALTH
UNIVERSITY OF MINNESOTA

Publishing Using SEER-Medicare Data

- Remember, the DUA requires that, PRIOR TO SUBMISSION, every publication must be submitted to NCI for approval
- Purpose is NOT peer-review
- Publications are reviewed for the following criteria:
  - Findings adhere to the SEER-Medicare data release policies.
  - Submitted publications are consistent with work proposed in the original application (focus on consistent research questions and measures)
  - Investigators do not report findings in which the cell size is less than eleven (CMS policy for cell size suppression)

Adhering to the Cell Suppression Policy for SEER-Medicare Policies

- No cell (e.g. admissions, discharges, patients, services, etc.) containing a value of 1 to 10 can be reported directly
- A value of *zero* does not violate the minimum cell size policy
- No cell can be reported that allows a value of 1 to 10 to be derived from other reported cells or information
- For SEER-Medicare data, applies to reporting of information on patients, providers and hospitals
- Purpose is to protect the confidentiality of Medicare and Medicaid beneficiaries and providers
- Examples can be found HERE

Final Thoughts

- We are happy to help!
- Submit any questions you have from requesting the data to understanding the SEER-Medicare files all the way to publication
- Questions can be submitted to the SEER-Medicare contact HERE

Questions?

Final Advice

Segment 7

SEER-Medicare Training 2019

Beth Virnig, PhD, MPH

Disclaimer

- I like to use SEER-Medicare data. It has allowed me to ask questions about cancer patients that would be difficult without the source.
  - How does pre-operative imaging relate to surgical decision-making?
  - Have specific policy or payment changes led to changes in treatment decisions and (ultimately) outcomes?
  - How does adherence to quality guidelines vary by patient and providers?

Disclaimer (part 2)

- I do not use SEER-Medicare data for every cancer question I have.
- I begin with a question and then find a data source
- I find this to be a more successful approach than beginning with a data source and trying to find a question.

Is SEER-Medicare the Right Data for Me?

- There is no perfect dataset, everything involves balancing strengths and weaknesses.
- Decision about whether to S-M should be made study by study
- S-M may be the right dataset for your question if:
  - You need to study incident cancers and need to know about care received, location of care, demographics, need a longitudinal perspective…
- S-M is NOT the right dataset for your study if:
  - You need information about health behaviors, height or weight, symptoms, margin status, results of clinical tests, information about care offered, work history, etc. (in other words, things not included in the data)
  - You want to study people under age 65

Is it Hard to Use?

- The data are complex:
  - You will need to combine multiple files
  - One-to-many or none structure
  - Variables will need to be created
- I use a “3 paper” rule
- Not good if you’re in a rush

How Do I Get Started

- Start with the PEDSF
- Add hospitalizations next
- Focus on something that has few(er) treatment options
- Use a narrower treatment window

What Else Do I Worry About:

- Will I have sufficient power?
- 116,535 breast cancer patients diagnosed between 2004 and 2013
- 94,024 (81%) of these had early stage (0-2) breast cancer
- 50,694 (44%) were treated with radiation therapy within a year after diagnosis
- BUT, only 5,408 (5%) also had a mastectomy. So, studying RT after mastectomy will be much harder than RT after breast conserving surgery…

Can I Find the Right Information to Address Clinical Uncertainty?

- Can I differentiate (in a clinically meaningful way) between clinically important categories such as high risk and low risk?
  - Adenocarcinoma—yes
  - Cancer among non-smokers--no

Timeliness of the Data

- SEER cancer registries are (by policy) given 2 years to complete identification of incident cancers and data abstraction
- Linkages are done every 2 years and include claims through the date of linkage
- At least a year is needed for data compiling, quality checks, etc.
- The current data were released in 2019 and are for incident cases from 2014-2015 with claims through 2017

Summary

- SEER-Medicare data are not easy, but are a tremendous resource. They are continually improving and offer amazing opportunities for understanding cancer care and outcomes!

Questions?

Q&A/Wrap Up