

*Training Sessions 3 and 4: Introduction to SEER,
Medicare and SEER-Medicare*

Rodney Sparapani, PhD
The Center for Patient Care and Outcomes Research
and the Division of Biostatistics
Medical College of Wisconsin

April 20 and May 4, 2012

Training Outline

1. **What is SEER?**
2. **What is Medicare?**
3. **What is SEER-Medicare?**
4. **Comorbidity via Medicare Claims**
5. **Breast Cancer Identification via Medicare Claims**
6. **Conclusions**

If all else fails, read the instructions.

- Donald Knuth, renowned computer scientist

Surveillance Epidemiology and End Results (SEER)

- **Free download** <http://seer.cancer.gov>
- **Docs** <http://seer.cancer.gov/analysis>
- **Ries, Pollack and Young. Cancer patient survival: Surveillance, Epidemiology, and End Results program, 1973-79. Journal of the National Cancer Institute 1983;70;693-707**
- **all incident cancers in certain registry areas (except non-melanoma skin cancer): 1973-2009**
- **de-identified, i.e. no names, SSN, etc.**
- **demographics, Census 2000, mortality/cause of death, and up to 10 lifetime primary cancer episodes**
- **cancer info: diagnostic/prognostic, initial treatment**
but not chemotherapy nor endocrine therapy

Surveillance Epidemiology and End Results (SEER) (cont.)



Medicare

- **US national health insurance program for the disabled and elderly (65+) managed by the Centers for Medicare and Medicaid Services (CMS)**
- **Standard Analytic Files (SAF) created for research**
- **Part A: claims for hospital/inpatient stays (MEDPAR/Inpatient SAF: ICD-9-CM diagnoses and procedures)**
- **Part B: claims for physician/clinic visits (Carrier and Outpatient SAFs: ICD-9-CM diagnoses and HCPCS procedures)**
- **Part C: claims for Medicare Advantage AKA HMO (A/B claims NA but D are available)**
- **Part D: claims for prescription drugs (PDE SAF: NDC, generic and brand names)**

Medicare and the Elderly Aged 65+ (circa 2008-2009)

- **Part A: 98% enrolled; in general, no monthly premium**
- **Part B: 94% enrolled; monthly premium**
- **Part C: 20% enrolled; covers Part A/B/D services in a managed care setting**
- **Part D: 58% enrolled (includes Part C: PDE claims are available)**
 - 17% enrolled in Retiree Drug Subsidy (claims NA)**
 - 14% not enrolled, but have Creditable Coverage (claims NA: Part D equivalent coverage)**
- Total: 89% have drug coverage**
- **Neuman and Cubanski. Medicare Part D Update: Lessons Learned and Unfinished Business. NEJM 2009;361;406-14.**

Medicare Demographics and Eligibility/Enrollment

- **Beneficiary Summary File:** `CCW.DENOM`
- **A/B/C/D eligibility/enrollment information**
- **Birth date and death date, if any**
- **Age:** `age=%ageyr (birth_dt, '01JAN2008'd);`
- **Gender**
- **Race (B/W/O/API/H/NA)** `format race race.;`
- **Research Triangle Institute (RTI) Race Imputation**
- **RTI Race** `http://www.ahrq.gov/qual/medicareindicators/medicareindicators2.htm`
- **Address, City, County, State and ZIP code**

Medicare Part A

- **Hospital Stays, 2 flavors: MEDPAR and Inpatient SAF (CCW . IP)**
- **Also Skilled Nursing Facility (SNF), Home Health Agency (HHA) and Hospice**
- **Inpatient SAF has more info than MEDPAR: mainly physician ID**
- **Both have Hospital ID, dates, charges, etc.**
- **Hospital Stays have up to 10 ICD-9-CM diagnosis and 6 ICD-9-CM procedure codes**

Medicare Part B

- **2 flavors: Carrier SAF and Outpatient SAF:** CCW.CO
- **Carrier SAF is for physician billing:** CCW.CARRIER
- **Outpatient SAF is for clinic billing:** CCW.OP
- **up to 10 ICD-9-CM diagnosis codes and 1 HCPCS procedure code**
- **Both have physician ID (unfortunately, only Outpatient has hospital ID while Carrier are more numerous)**
- **UPIN used up to June 2007; replaced by NPI** <https://npes.cms.hhs.gov/NPPES/NPIRegistryHome.do>
- **neither are unique; only the AMA ResearchID is unique**
- **with approval for research, you can buy from the AMA** <http://www.mmslists.com>
with a ResearchID to UPIN/NPI crosswalk

Medicare Part C

- **Medicare Advantage AKA Health Maintenance Organization (HMO)**
- **Part A and B claims appear in their respective places**
- **But, HMOs do not provide all of their claims to CMS since there is no additional compensation**
- **Therefore, we often omit Part C patients from our cohorts since their treatments and services are incomplete**
- **However, all Part D claims are provided to CMS**

Medicare Part D

- **Part D Event SAF:** CCW . PDE
- **Enrollment began January 2006, but not compulsory until June 2006**
- **National Drug Code (NDC), generic/compound name, brand name**
- **fill date, pill count, charges, etc.**
- **Part C enrollee claims available**
- **Part D Retiree Drug Subsidy claims not available**
- **Non-part D Creditable Coverage claims not available**

Medicare and HIPAA

- **Health Insurance Portability and Accountability Act (HIPAA) <http://infoscope.mcw.edu/hipaa.htm>**
- **patients are identified, i.e. names, SSN, etc.**
- **cannot easily be de-identified, i.e. Medicare includes the dates served by, and the identifiers of, hospitals, clinics, physicians, etc.**
- **Research practices and security dictated by Data Use Agreement (DUA)**
- **HIPAA not applicable to Medicare claims for research purposes, but DUA has similar requirements**
- **PCOR SOP 002: Good Protected Health Information Practices**

Medicare Resources

- **Medicare**
`http://www.cms.gov/home/medicare.asp`
- **Chronic Condition Warehouse (CCW) Documentation**
`http://www.ccwdata.org/data-dictionaries`
- **Research Data Assistance Center (ResDAC)
Medicare Documentation**
`http://www.resdac.org/ddvh/Index.asp`
- **SES Part D Study:** `/ses/doc/Medicare`
- **Diagnosis/Procedure codes (except CPT):**
`/ses/doc/Medicare/Code Reference Sets_2008.xls`

Medicare Coding Resources

- **International Classification of Disease, ICD-9-CM:**
<http://www.cdc.gov/nchs/icd/icd9cm.htm>
- **ICD-10-CM:**
<http://www.cdc.gov/nchs/icd/icd10cm.htm>
- **American Medical Association (AMA) Current Procedural Terminology (CPT) AKA HCPCS Level I: code books in MCW library and PCOR**
- **Health Care Procedure Coding System HCPCS Level II:**
https://www.cms.gov/HCPCSReleaseCodeSets/02_HCPCS_Quarterly_Update.asp
- **National Drug Code (NDC):**
<http://www.accessdata.fda.gov/scripts/cder/ndc/default.cfm>

SEER-Medicare

- **http:**
`//healthservices.cancer.gov/seermedicare`
- **combines SEER (1973-2007) with Medicare claims (1991-2009; Part D 2007-2009): adds treatment information like chemotherapy and endocrine therapy**
- **we focus on SEER (2000-2007) with Medicare claims (1999-2009)**
- **includes a 5% Medicare non-cancer control group**
- **de-identified: no names, SSN, address, etc.**
- **no other identifiers like hospital, clinic, physician, etc.**
- **dates of service are available**
- **Patient Entitlement and Diagnosis Summary File (PEDSF) /seermedicare/doc/new/PEDSF.pdf**

SEER-Medicare and Breast Cancer (ICD-10: C50)

- **We focus on the patient's first lifetime cancer, i.e. `dx_dt1` and nearly all other variables that end in 1 (created variables `brstins`, `brstmal` and `erpr` do not)**
- **Patients who's first lifetime cancer is a breast cancer are identified as follows:**
 - **in situ (Stage 0) breast cancer** `brstins=1`
 - **invasive (Stage 1-4) breast cancer** `brstmal=1`
- **Diagnostic/prognostic variables:**
 - **Extension of Disease (EOD): TNM Stage Classification**
 - **Tumor Grade** `grade1`
 - **Tumor Markers: Estrogen Receptor (ER) and Progesterone Receptor (PR)**
 - **ER and PR combined into one variable** `erpr`

SEER-Medicare and Extension of Disease (EOD)

- **American Joint Committee on Cancer (AJCC) staging:**
0 (in situ), 1-4 (invasive) and 4 is metastatic
- **TNM Stage Classification:** T for primary Tumor, N for regional lymph Nodes and M for distant Metastasis
- **4 variables:** T, tumor size (e10sz1);
N, nodal involvement (e10nd1) and positive nodes (e10pn1); and M, tumor extension (e10ex1)
- **Generally, the TNM components are 3 separate variables (T, N and M) rather than Stage (0-4) due to missingness (TX, NX and MX) i.e. if the regional lymph nodes were not removed, then N is unknown (NX)**
- **For breast cancer, page 120 of <http://seer.cancer.gov/manuals/EOD10Dig.pub.pdf>
[/seermedicare/doc/new/EOD10Dig.pub.pdf](http://seermedicare/doc/new/EOD10Dig.pub.pdf)**

SEER and Missing Values

- **SEER variables are somewhat consistently coded for missingness**
- **For example, ER is 0 for 'Not Done', 3 for 'Borderline', 8 for 'Not in Chart' and 9 for 'Unknown'**
- **For our purposes, these have all been recoded to a SAS missing value `._`**
- **See `/seermedicare/sas/new/pedsf.sas` which creates `BREAST.PEDSF`**

SEER and Missing Values with Informats

```
*example based on /seermedicare/sas/new/pedsf.sas;
proc format;
    invalue _0389_
        0, 3, 8, 9=._
        2          =0; *1 does not need re-coding;
run;
data erpr;
input ... tumor1_1 _0389_1. tumor2_1 _0389_1. ...;
...
    if brstmal=1 | brstins=1 then do;
        if tumor1_1 | tumor2_1 then erpr=1;
        else if tumor1_1=0 & tumor2_1=0 then erpr=0;
    end;
run;
```

Comorbidity

- **Comorbid conditions: conditions unrelated to the disease being studied**
- **We follow the Charlson approach:**
Charlson et al. J Chron Dis 1987;40:373-83
- **With “enhanced” ICD diagnoses (see**
*/ses/sas/ccw/enhanced.sas***):**
Quan et al. Medical Care 2005;43:1130-9
- **NCI Combined: combining inpatient and outpatient claims for the 4 major cancers; breast, colorectal, lung and prostate**
Klabunde et al. Ann Epidemiol 2007;17:584-90

Comorbidity Warehouses and SAS Programs

- We define comorbidities in time “windows”
- e.g. during the period starting one year prior to surgery up to the day before surgery
- OR from two years prior until surgery
- To facilitate these varying definitions: we construct static warehouses which need to re-built only when new Medicare data arrives (rarely)
- **Inpatient:** `inmorbidity.sas` **creates** `CCW.INMORBID`
- **Outpatient:** `outmorbidity.sas` **creates** `CCW.OUTMORBID`

Comorbidity Warehouses and SAS Programs

- CCW.INMORBID **and** CCW.OUTMORBID
- **Non-cancer Diagnosis/Disease: 1. Acute MI, 2. Congestive Heart, 3. Peripheral Vascular, 4. Cerebrovascular, 5. Dementia, 6. Chronic Pulmonary, 7. Rheumatologic, 8. Peptic Ulcer, 9. Mild Liver, 10. Mild/moderate Diabetes, 11. Severe Diabetes, 12. Hemiplegia/Paraplegia, 13. Renal, 14. Moderate/Severe Liver, 15. AIDS, 16. Old MI**

Comorbidity Warehouses and SAS Programs

- CCW.INMORBID **and** CCW.OUTMORBID
- **Primary and Secondary Neoplasms**
- **System: 20. Oral/Pharynx, 21. Digestive, 22. Respiratory, 23. Bones/Joints, 24. Soft Tissue, 25. Melanoma, 26. Breast, 27. Genitals, 28. Leukemia, 29. Urinary, 30. Eye/Orbit, 31. Brain/Nervous, 32. Endocrine, 33. Lymphoma/Myeloma, 34. Misc.**

Comorbidity Warehouses and SAS Programs: Bit Arithmetic

- **Combined:** `comorbid.sas` **creates** `CCW.COMORBID` **from** `CCW.INMORBID` **and** `CCW.OUTMORBID`
- **Example 1: No Primary Breast Cancer diagnoses in June 2006**
- `P20060626=0`
- **Example 2: Primary Breast Cancer diagnoses in June 2006 on the 1st and 15th**
- `P20060626=32770 = 21 + 215`
- `band(2**15, P20060626) = 32768 = 215`
- `band(2**10, P20060626) = 0`

Medicare, Breast Cancer and Comorbidity Warehouses

- **Combined:** `charlson.sas` summarizes `CCW.INMORBID` and `CCW.OUTMORBID`:
`CCW.CHARLSON12` requires 12 months of prior claims and `CCW.CHARLSON24` requires 12 months of prior claims and 12 months of “confirmation” claims
- **16 Non-cancer Diagnoses:** `C20000101--C20081216`
- **15 Primary Cancers:** `P20000120--P20081234`
- **15 Secondary Cancers:** `S20000120--S20081234`
- **Also:** `/ses/sas/ccw/alg.sas` summarizes `CCW.INMORBID` and `CCW.OUTMORBID` for identifying breast cancer patients (but that also needs surgeries)
- **Once you have identified breast cancer patients: you can calculate their non-cancer comorbidities via NCI Combined**

Medicare, Breast Cancer and Comorbidity Warehouses

```
merge
ccw.charlson12(keep=bene_id c20010101--c20081216)
...;
by bene_id;

%_array(array=_c, dim1=2001:2008, dim2=12, dim3=16,
length=3, var=c);

array _cwgt(16) _temporary_
(0.201 0.845 0.224 0.713 1.192 0.471 0.751 0.228
0      0.450 0.023 0.210 1.188 0      0      0.396);

if nmiss(_c(dx_yr1, dx_mol, 1)) then nci=-1;
```

Medicare, Breast Cancer and Comorbidity Warehouses

```
else do;
  if esrd then _c(dx_yr1, dx_mo1, 13)=1;

  if _c(dx_yr1, dx_mo1, 11)
  then _c(dx_yr1, dx_mo1, 10)=1;

  if _c(dx_yr1, dx_mo1, 14)
  then _c(dx_yr1, dx_mo1, 9)=1;

  nci=0;
  do i=1 to 16;
    if _c(dx_yr1, dx_mo1, i)>0
    then nci=(nci+_cwgt(i));
  end;
end;
```

Identifying Breast Cancer with Medicare

First Author	Year	Sens.	Spec.	PPV	Medicare Claims
McClish	1997	83.0%			Hospital only
Cooper	1999	82.0% 93.6%			Hospital only Hospital/Physician
Warren	1999	57.0% 76.2%	99.9% 99.3%	91.3% 36.3%	Hospital only Hospital/Physician
Freeman	2000	90.0%	99.86%	70.0%	Hospital, Physician & Clinic
Nattinger	2004	80.1%	99.95%	89.0%	Hospital, Physician & Clinic

Identifying Breast Cancer with Medicare

Study	PPV	95% CI	Prior Year(s)	Cohort
Nattinger et al.	89.0%	(86.7%, 91.6%)	3	Validation Training
	88.1%	(85.6%, 90.7%)	3	
Survey	82.5%		1	CA
	86.9%		4	FL/IL/NY
	85.7%		3.2-3.25	All

The Breast Cancer Identification Algorithm for Medicare Claims

- S1.** negative if either no breast cancer diagnosis OR no breast cancer procedure; otherwise S2
- S2.** Both needed to be positive, go to S4; otherwise S3
- mastectomy OR {a lumpectomy/partial mastectomy AND a radiotherapy claim with a breast cancer diagnosis}
 - 2+ claims 31+ days apart with a primary diagnosis of breast cancer
- S3.** All 3 needed to be positive, go to S4; otherwise negative
- a mastectomy OR a lumpectomy/partial mastectomy
 - 2+ claims 31+ days apart with a diagnosis of primary or in situ breast cancer
 - neither a secondary diagnosis of breast cancer nor a primary diagnosis of another cancer
- S4.** negative if in prior 3 years: a history of breast cancer diagnosis OR {a breast cancer diagnosis AND a lumpectomy/partial mastectomy/mastectomy}

The Breast Cancer Identification Algorithm for Medicare Claims
(*cont.*)

Nattinger, Laud, Bajorunaite, Sparapani, Freeman.
An algorithm for the use of Medicare claims data to identify
women with incident breast cancer.
HSR 2004:39;1733-49

Here Tables 6 and 7 replace the original Table 1

The Breast Cancer Identification Algorithm for Medicare Claims

(cont.)

Table 6. Diagnosis Codes (SIMPLIFIED)

Diagnosis (Algorithm Step(s) Involved)	ICD-9-CM Diagnosis Code(s)
Primary breast cancer (1, 2, 4)	174.x
Breast carcinoma in situ (1, 2)	233.0
Secondary breast cancer (3)	198.2, 198.81
Non-breast cancer (3)	140.x-208.91, 230.x-239.9 (except 174.x, 196.x, 198.2, 198.81, 232.5, 233.0, 238.3, 239.3)
Breast cancer history (4)	V10.3

The Breast Cancer Identification Algorithm for Medicare Claims

The Breast Cancer Identification Algorithm for Medicare Claims

(cont.)

Table 7. Procedure Codes (CORRECTED AND EXPANDED)

Procedures (Algorithm Step(s), if any)	ICD-9-CM Procedure Code(s)	CPT/HCPCS Code(s)
Tumor Biopsy (1)	85.1x	19000, 19001, 19100-19103, 19110, 19112
Lymph node biopsy/dissection only (1)	40.23 (axillary), 40.3 (regional), 40.50 (unspecified), 40.51 (axillary)	38500 (axillary), 38525 (axillary), 38740 (axillary), 38745 (axillary)
Lumpectomy (1-4)	85.20, 85.21	19120, 19125, 19126
Partial mastectomy (1-4)	85.22, 85.23	19160, 19162, 19301, 19302
Mastectomy (1-4)	85.33-85.36, 85.41-85.48	19180, 19182, 19200, 19220, 19240, 19303-19307
Radiotherapy (2)		77401-77404, 77406-77409, 77411-77414, 77416-77418, 77427, 77431, 77520, 77522, 77523, 77525, 77750, 77761, 77763, 77776-77778, 77781-77784
Chemotherapy		J8510, J8520, J8521, J8530-J9999 (except J9202, J9209, J9212-J9214, J9217, J9218)
Lymph node dissection with other procedure	85.43-85.48 (axillary)	19162 (axillary), 19200 (axillary), 19220 (axillary & mammary), 19240 (axillary), 19302 (axillary), 19305 (axillary), 19306 (axillary & mammary), 19307 (axillary)
Sentinel lymph node biopsy		A9520 (radioactive dye, 2003-2005), A9541 (radioactive dye, 2006+), 38500 (axillary), 38525 (axillary), 38790 (injection), 38792 (injection)

Outcomes of older women with early stage breast cancer

- **Updated Epi/PCOR presentation written by Rodney; generally, you can trust it, even when it conflicts with a publication: /survey/doc/survey.pdf**
- **Nattinger A, Pezzin L, Sparapani R, Neuner J, King T, Laud P. Heightened attention to medical privacy: challenges for unbiased sample recruitment and a possible solution. American Journal of Epidemiology 2010;172(6):637-44. PMC3025645**
- **Neuner J, Yen T, Sparapani R, Laud P, Nattinger A. Fracture risk and adjuvant hormonal therapy among a population-based cohort of older female breast cancer patients. Osteoporosis International 2011;22:2847-55. PMC3166362**

Outcomes of older women with early stage breast cancer (cont.)

- **Kong A, Yen T, Pezzin L, Miao H, Sparapani R, Laud P, Nattinger A. Socioeconomic and racial differences in treatment for breast cancer at a low-volume hospital. Annals of Surgical Oncology 2011;18:3220-7. PMC3201787**
- **Yen T, Czypinski L, Sparapani R, Guo C, Laud P, Pezzin L, Nattinger A. Socioeconomic factors associated with adjuvant hormone therapy use in older breast cancer survivors. Cancer 2011;117(2):398-405. PMC3010527**
- **Yen T, Sparapani R, Guo C, Neuner J, Laud P, Nattinger A. Elderly breast cancer survivors accurately self-report key treatment information. Journal of the American Geriatrics Society 2010;58(2):410-2. PMC2972547**

Outcomes of older women with early stage breast cancer (cont.)

- **Yen T, Fan X, Sparapani R, Laud P, Walker A, Nattinger A. A contemporary, population-based study of lymphedema risk factors in older women with breast cancer. *Annals of Surgical Oncology* 2009;16(4):979-88. PMC2729500**