

7th Annual Epidemiology, Biostatistics and Clinical
Research Methods Summer Session
June 20-24, 2005

Using VA Data with Non-VA Data

Session Objectives

- Know about the VA-Medicare Data Merge Initiative and available data
- Understand how to request VA-Medicare data
- Know where to go for help
- Know about data available from NCI-SEER linked Medicare data
- Know about the VA Central Cancer Registry

Session Objectives

- **Know about the VA-Medicare Data Merge Initiative and available data**
- Understand how to request VA-Medicare data
- Know where to go for help
- Know about data available from NCI-SEER linked Medicare data
- Know about the VA Central Cancer Registry

VA Medicare Data

Why do you need Medicare data?

- More than 40% VHA users 65 years and older also use Medicare services
 - Hynes, et al., VA USH Report, 2003
- Prevalence of diseases of VHA users underestimated without Medicare data
 - Tseng et al., Am J Managed Care, Nov 2004
- Complications not correctly detected without Medicare data
 - Fleming & Fisher, Med Care, 1992
- Risk scores may not be correctly estimated without Medicare data

VA-Medicare Data Merge Initiative

- Centers for Medicare and Medicaid Services (CMS) & VHA Interagency Agreement
- Medicare data since 1999 are available through VHA
 - For researchers through VIREC
 - For all other purposes through the ADUSH Office
- Linkable with VA data
- All veterans “known to the VA” since 1997
- Data available to VA researchers with IRB approved research

Veterans Included

■ Includes:

- VHA users
- Enrollees
- C&P Beneficiaries



■ Veterans Identified Increasing

- 1999 – 6.4 million
- 2003 – 12.6 million

Process of Creating Linkable Files

- Identify veterans from VA sources
- Obtain VA demographic information
- Send SSN list to CMS
- Receive cross-walk of SSN and Health Insurance Claim (HIC) Account Numbers from CMS with demographic information from CMS
- VIREC receives data, verifies the data, and attaches real and scrambled SSNs

Linkage Results

Service Year	Finder File SSNs	Denominator SSNs	Cumulative Vital Status SSNs
1999	6,411,210	3,242,544	3,437,360
2000	9,717,490	3,820,989	N/A
2001	10,646,812	4,003,734	N/A
2002	12,032,892	4,541,142	6,151,668
2003	12,614,460	4,653,805	6,543,766

Medicare Data Available

- CY1999 – CY2003
- Enrollment/Demographic: Denominator, Vital Status
- Claims
 - Standard Analytical Files (SAFs)
 - Summary File (MedPAR)
- Other Medicare files
 - Provider of Services
 - Unique Physician Identification Number (UPIN)
 - 100% Denominator
 - Medicare Current Beneficiary Survey

Medicare Enrollment & Demographic Files Denominator File

- All beneficiaries enrolled and/or entitled in a given year
- One record per beneficiary
- Content frozen in March of the following year
- Demographic characteristics
 - State, County, Zip Codes
 - Date of Birth, Date of Death, Sex, Race

Medicare Enrollment & Demographic Data Denominator File

- Entitlement indicators
 - Reasons for entitlement
 - State buy-in indicators
 - Termination indicators and reason
- Managed care enrollment indicators
 - Monthly flags

Percent Enrolled in Medicare+Choice By Number of Months

Enrolled	1999	2000	2001	2002	2003
0 Months	83.0%	82.5%	84.5%	86.6%	87.6%
< 12 Months	3.2%	3.0%	3.3%	2.4%	2.1%
12 Months	13.8%	14.5%	12.2%	11.0%	10.4%

Note: Over 12 % receiving care under Medicare HMO - there are no claims data to describe that care

Medicare Enrollment & Demographic Files Vital Status File

- Ever enrolled in Medicare (cumulative)
- State and county codes, ZIP code
- Name, sex, race
- Date of birth and death
- Death information current when file created

Medicare Standard Analytical Files (SAFs)

- Contain final action claims
- Diagnosis (ICD-9-CM)
- Procedure (ICD-9 & HCPCS)
 - ICD-9 for Inpatient, SNF, and Hospice SAFs
 - HCPCS for Outpatient, Carrier, and DME SAFs
- Dates of service
- Charge and reimbursement amount – Revenue center
- Provider number
- Beneficiary demographics: DOB, sex, ZIP code

Medicare SAFs

- Institutional (~2,400 variables)
 - Inpatient
 - Outpatient
 - Skilled Nursing Facility
 - Home Health Agency
 - Hospice
- Non-institutional (~1,000 variables)
 - Physician/Supplier (Carrier)
 - Durable Medical Equipment

Medicare Institutional: Inpatient SAF

- Claims by hospitals for inpatient care
- Reimbursement of facility costs
- Admission & discharge date
- Revenue center detail

Medicare Institutional: Skilled Nursing Facility SAF

- Claims by Skilled Nursing Facilities (SNF)
- Variables similar to Inpatient SAF
- Revenue center detail
- Combined with Inpatient SAF in MedPAR

Medicare Institutional: Outpatient SAF

- Claims by institutional outpatient providers:
 - Hospital outpatient departments
 - Rural health clinics
 - Renal dialysis facilities
 - Outpatient rehabilitation facilities
 - Community mental health centers
 - Ambulatory surgical centers

Medicare Institutional: HHA SAF

- Claims by Home Health Agencies (HHA)
- Type of visit
 - Skilled-Nursing Care
 - Home Health Aide
 - Physical Therapy
 - Occupational Therapy
 - Speech Therapy
 - Medical Social Services

Medicare Institutional: Hospice SAF

- Level of hospice care received
 - Routine home care
 - Inpatient respite care
- Terminal diagnosis (ICD-9-CM)

Medicare Non-Institutional: Physician/Supplier SAF

- AKA “Carrier SAF”
- Claims by Non-institutional providers
 - Physicians
 - Physician assistants
 - Clinical social workers
 - Nurse practitioners
 - Independent clinical laboratories
 - Ambulance providers
 - Stand-alone ambulatory surgical centers
- CMS’ Healthcare Common Procedure Coding System (HCPCS) codes

Medicare Non-Institutional: Durable Medical Equipment (DME) SAF

- Claims by DME Suppliers
- Ordering Physician UPIN Number
- Performing Physician UPIN Number
- Referring Physician UPIN Number
- Provider Specialty
- Type of Service

Medicare Summary File: Medicare Provider Analysis and Review

- Inpatient - discharge summary
- Skilled Nursing – may “still be a patient”
- One record per stay
- Revenue center totals

Other Medicare Datasets: Provider of Services File

- Contains data on institutional providers
- Institutional providers
 - Hospital
 - Skilled Nursing Facility
 - All other institutional providers
- Linkable to claims files using Medicare Provider Number
- Name, address, and facility characteristics

Other Medicare Datasets:

UPIN File

- Contains data on non-institutional providers
 - physicians
 - Licensed practitioners
 - Non-physician practitioners
- Unique Physician Identification Number (UPIN)
- Linkable to claims files using UPIN
- Name, specialty, license code, and state
- Available for CY 2001 – CY 2003

Other Medicare Datasets: 100% Denominator File

- Cross-reference for veteran cohort
- 43.1 million individuals in CY2003
- 4.7 million veterans (10.8%) in CY2003
- De-identified data for non-veterans

Other Medicare Datasets: Medicare Current Beneficiary Survey (MCBS)

- Cross-reference for veteran cohort
- 16,315 individuals in CY2002
- 1,838 (11.3%) veterans in CY2002
- Access to Care File
- Cost and Use File
- For more information go to:
<http://www.resdac.umn.edu/MCBS/Index.asp>

Session Learning Objectives

- To learn about the VA-Medicare Data Merge Initiative and available data
- **How to request VA-Medicare data**
- Where to go for help
- To learn about data available from NCI-SEER linked Medicare data
- To learn about the VA Central Cancer Registry

Requesting Medicare Data

- VA researchers with IRB approved research projects can request data from VIREC
- Documents needed:
 - Signed data request form
 - Local VA R&D approval letter
 - IRB approved research protocol
 - IRB approval letter
 - IRB approved data security plan
 - IRB waiver of informed consent
 - HIPAA waiver of authorization
- Information to request data is available on the Web:
<http://www.virec.research.med.va.gov/DataSourcesName/VA-MedicareData/DataRequest/Procedure.htm>

Requesting Medicare Data

- For more information, contact VIREC staff at virec_medicare@va.gov
- Or call (708) 202-2413

Session Learning Objectives

- To learn about the VA-Medicare Data Merge Initiative and available data
- How to request VA-Medicare data
- **Where to go for help**
- To learn about data available from NCI-SEER linked Medicare data
- To learn about the VA Central Cancer Registry

Documentation & Help

- VIREC Website

- <http://www.virec.research.med.va.gov/>
- Updated as files are available for release
- Links to CMS & ResDAC

- *VIREC Insights:*

- VA-Medicare Data for Research
- <http://www.virec.research.med.va.gov/References/VirecInsights/Insights-v04n3.pdf>

- USH Report: available on VIREC Web site

- <http://www.virec.research.med.va.gov/DataSourcesName/VA-MedicareData/USHreport.pdf>

**Data Sources by Name**

- ▶ BIRLS Death File
- ▶ DSS
- ▶ Medical SAS Datasets
- ▶ NPCD
- ▶ NPPD  PDF
- ▶ PBM
- ▶ VA Medicare Data
- ▶ VistA

Data Sources by Category**Help****Library****About VIREC**

VA MEDICARE DATA



FOR RESEARCH

Available VA-Medicare Data

Contents of VA Medicare Data Pages

- [Available VA Medicare Datasets](#)
- [VA Medicare Data Initiative Overview](#)
- [Cohort](#)
- [How to Request Data](#)
- [FAQs](#)
- [Help](#)
- [USH Report](#) ( PDF 7,095KB)

Description

Medicare data for veterans are available to VA researchers with approved projects. VIREC prepares the data and attaches it to VA's Scrambled Social Security Numbers, which allows researchers to link VA and Medicare healthcare utilization data. The data are distributed to VA researchers as SAS datasets. See [VIREC Insights Vol 4, No 2](#)



Veterans Affairs Information Resource Center

REQUEST FOR PATIENT LEVEL MEDICARE DATA

PRIVACY STATEMENT: It is the policy of VIREC to protect the patient's rights of confidentiality. The requestor, in exchange for receipt of patient level data, agrees to use the data only as described and for the purpose(s) set forth in this request. The requestor further agrees to provide a secure environment at a VA facility for storage and use of the source data and any working files to prevent unauthorized access. Except as specified in the attached request, or as later approved, the requestor agrees not to release, share or further distribute these records and not to release, share or distribute any data containing complete or partial patient records. The requestor will comply with all laws, regulations and VA/VHA policies relating to privacy of patient information.

DATA RETENTION POLICY: While the research project is active, the requestor agrees to submit a copy of the annual IRB renewal letter and any IRB amendments related to Medicare data. VA-Medicare data from VIREC may be retained for as long as the project maintains IRB approval. When IRB approval for a project ends, the requestor must arrange with VIREC the return or destruction of all datasets containing protected health information.

REQUESTOR INFORMATION

	PRINCIPAL INVESTIGATOR	CONTACT PERSON (optional)
A. NAME		
B. TITLE		
C. ORGANIZATION		
D. VA SERVICE		
E. VA FACILITY NAME		
F. VA STATION NUMBER		
G. MAILING ADDRESS		
H. CITY, STATE, ZIP		
I. PHONE NUMBER		

Research Data Assistance Center (ResDAC)

- CMS contractor based at the University of Minnesota with affiliates at Boston University, University of Dartmouth, & Morehouse University
- Provides free assistance to
 - Researchers
 - Government agencies
 - Not-for-profit organizations
- Website (<http://www.resdac.umn.edu>)
 - Record layouts
 - Codebooks
 - Links to other CMS websites

Questions?

Session Objectives

- Know about the VA-Medicare Data Merge Initiative and available data
- Understand how to request VA-Medicare data
- Know where to go for help
- **Know about data available from NCI-SEER**
- Know about the VA Central Cancer Registry

7th Annual Epidemiology, Biostatistics and Clinical
Research Methods Summer Session

June 20-24, 2005

**National Cancer Institute
(NCI)
Surveillance, Epidemiology,
and End Results
(SEER)
Program Data**

NCI-SEER

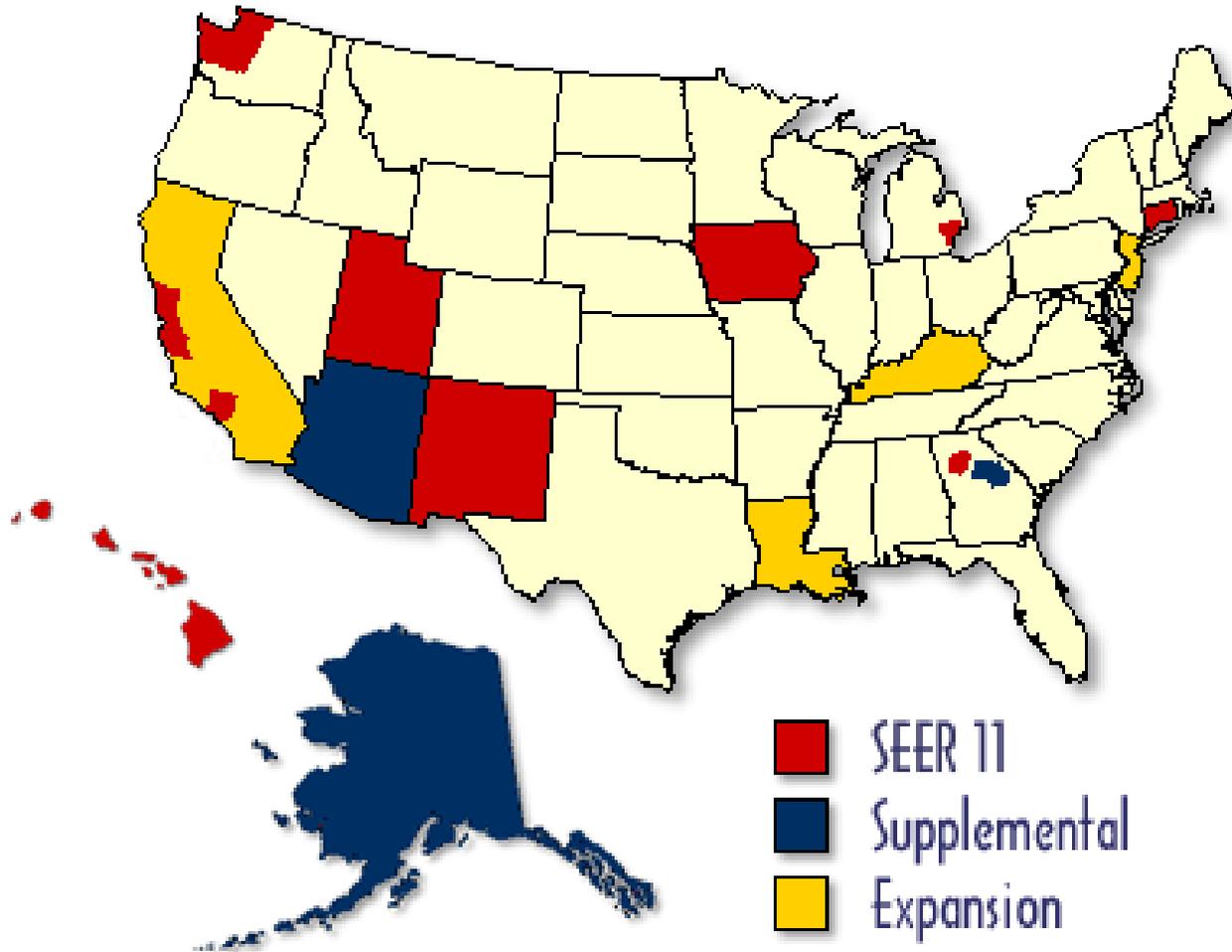
Surveillance, Epidemiology, and End Results Program

- Authoritative source of information on cancer incidence and survival in the US
- Began collecting data on cases on 1/1/1973
- Only comprehensive source of population-based information including stage at the time of diagnosis and survival rates within each stage

SEER Cancer Registries

- Includes 14 population-based cancer registries and three supplemental registries
- Covers approximately 26% of US population
- Information on 3 million *in situ* and invasive cancer cases in the SEER database
- About 170,000 new cases added each year

SEER Reporting Areas



Data Elements Collected

- Patient demographics
- Primary tumor site
- Morphology
- Stage at diagnosis
- First course of treatment
- Vital status

Data Access

- Public Use Data
 - Data Use Agreement
 - Three databases to request
 - SEER 9
 - SEER 13
 - SEER 13 without Alaska
 - No patient identifier

- Special data linkage at SEER sites possible

- More information on:
 - <http://seer.cancer.gov/publicdata/>



- SEER 1973-2002 Public-Use Data
- [Accessing the SEER Public-Use Data](#)
- [Access Options](#)
- [Submit Request for the Data](#)
- [Documentation for the Data Files](#)
- [Suggested Citations](#)
- Related Materials:**
- [SEER*Stat Software](#)
- Other Databases:**
- [US Mortality](#)
- [US Population Data](#)
- [Standard Population Data](#)

SEER 1973-2002 Public-Use Data

The SEER Public-Use Data include SEER incidence and population data associated by age, sex, race, year of diagnosis, and geographic areas (including [SEER registry](#) and county). A signed SEER Public-Use Data Agreement is required to [access these data](#). Use of these data for publication purposes should contain a citation which includes submission and release dates. See [Citations for SEER Public-Use Data and SEER*Stat Software](#) for specific suggestions regarding the citation format.

[Documentation for the SEER Public-Use Data](#)

SEER 9 Registry Database

The SEER 9 registries are Atlanta, Connecticut, Detroit, Hawaii, Iowa, New Mexico, San Francisco-Oakland, Seattle-Puget Sound, and Utah. In this data set, cases diagnosed from 1973 through 2002 are available for all registries except Seattle-Puget Sound (1974+) and Atlanta (1975+). The database contains one record for each of 3,122,042 tumors. Cases are associated to the population data using three racial groups: White, Black, and Other.

SEER 13 Registry Database

This database contains data from the SEER 9 registries (listed above), plus Los Angeles, San Jose-Monterey, Rural Georgia, and the Alaska Native Tumor Registry. In population-linked versions of these data, cases diagnosed from 1992 through 2002 are available, one record for each of 1,941,393 tumors. Cases are associated to the population data using four racial groups: White, Black, American Indian/Alaska Native, Asian/Pacific Islander. The Other race category used in the SEER 9 Registry database consists of American Indian/Alaska Native and Asian/Pacific Islander combined.

In SEER*Stat sessions that do not require population data (e.g., frequencies, survival, case listing), all public data for SEER registries are included in one database (SEER 13 Registry 1973-2002 with varying years of diagnosis per registry). This database includes 3,645,924 tumors diagnosed from 1973-2002 for the registries other than Seattle-Puget Sound (1974+), Atlanta (1975+), Los Angeles (1992+), San Jose-Monterey (1992+), Alaska Native Tumor Registry (1992+), and Rural Georgia (1992+).

Data Quality Assurance

- Standard for quality among cancer registries around the world
- Quality control has been an integral part of SEER since its inception
- Annual evaluation of quality and completeness of the data being reported in SEER sites

Completeness in SEER Data

- Underreporting may be a problem
- McClish et al. *Med Care* 2004
 - Registry captures
 - 59% of colorectal cancer cases
 - 74% of lung cancer cases
 - Registry & Medicare claims
 - 79% of colorectal cancer cases
 - 82% of lung cancer cases
- Supplementing cancer registry data with other sources
 - Cause of death in death certificates (Stang et al. 1999)
 - Hospital discharge data (Penberthy et al. 2002)

Other Data Quality Issues

- Concern with the completeness of treatment information (e.g., adjuvant therapies)
 - Cress et al. *Med Care*. 2003
 - 82% of chemotherapy
 - 90.2% of radiation therapy
 - Without utilization data, cancer registry data alone can lead to biased estimates in adjuvant therapies provided
- Concern with staging data
 - Fanning et al. 2000. *Gyn Oncol*. 2000

SEER Cancer Statistics Review

- An annual report with statistics on
 - Incidence
 - Mortality
 - Survival
 - Prevalence
 - Lifetime risk
- Latest report includes statistics from 1975 through 2002
- http://seer.cancer.gov/csr/1975_2002/

SEER-Medicare Linked Data

- Collaborative effort of NCI, the SEER registries and CMS
- Large population-based source of information for cancer-related epidemiologic and health services

SEER-Medicare Linkage Project

- SEER clinical data linked with Medicare claims data at NCI and CMS
- Patterns of care for persons with cancer
- Use of tests and procedures before and after cancer diagnosis
- Costs of cancer treatment
- Warren et al. *Med Care* 2002

SEER-Medicare Data Elements

- SEER demographic and clinical data

- Medicare
 - Enrollment data
 - All claims
 - Provider characteristics (POS, HCRIS)
 - Data for 5% cancer-free individuals in SEER areas

- Select variables from
 - Census data
 - Area Resource File

Cancer-Related Preventive Services: What's Included?

- Screening Mammography
 - Biennially for women 50+ starting in 1991
 - Annually for women 50+ starting in 1998
- Screening PSA
 - Annually starting in 2000

Cancer-Related Preventive Services: What's Included?

- Colorectal screening (all starting in 1998)
 - FOBT annually
 - Flex sig every 4 years for beneficiaries 50+
 - Colonoscopy every 2 years for high risk persons
- Pap Smear and Pelvic
 - Every 3 years, if normal, starting in 1998

Services Not Captured: What's Missing?

- HMO enrollees
- Care provided in other settings, such as the VA
- Care for persons with Medicare as the secondary payer

Reimbursement for Covered Services Not Captured

- Out of pocket expenditures
- Coverage provided by Medigap policies

Use of Radiation Therapy or Chemotherapy

■ Radiation Therapy

- ICD Diagnosis: V58.0, V66,1, V67.1
- ICD Procedure: 92.2-92.29, inclusive
- HCPCS: 77400-7749;77750-77797
- Revenue Center Codes: 0330, 0333

Use of Radiation Therapy or Chemotherapy

■ Chemotherapy

- HCPCS: 964xx, 965xx, J9000-9999, Q0083-Q0085 (other J codes have information about the administration of other drugs)
- ICD Diagnosis: V58.1, V66.2, V67.2
- ICD Procedure: 99.25

More Caveats

- ICD-9 codes and HCPCS change over time
- Longitudinal studies should include codes that were current with the study year

Resources for SEER-Medicare

- SEER-Medicare site on the Web
 - <http://healthservices.cancer.gov/seermedicare/>
- ResDAC Course on SEER-Medicare Data
 - <http://www.resdac.umn.edu/training/CMS301.asp>
- *Medical Care* Vol. 40, no. 8 August 2002
Supplement

Session Objectives

- Know about the VA-Medicare Data Merge Initiative and available data
- Understand how to request VA-Medicare data
- Know where to go for help
- Know about data available from NCI-SEER linked Medicare data
- **Know about the VA Central Cancer Registry**

VA Central Cancer Registry (VACCR)

- Administered by the Chief, Program Office for Oncology at VACO Washington DC
- Implemented under VHA Directive 97-050 *National Cancer Strategy* in FY98
- Aggregates data from local registries
 - 120 VA Medical Centers with tumor registries

VA CCR Coding Standards Comparable to SEER

- American College of Surgeon's Registry Operations And Data Standards ([ROADS](#))
- the NCI's [Surveillance, Epidemiology and End Results Program \(SEER\)](#)
- American Joint Commission on Cancer's [Manual for the Staging of Cancer](#)
- [International Classification of Diseases for Oncology \(ICD-O-2/3\)](#)

VACCR Data Elements

- Patient demographics
- Primary tumor site
- Morphology
- Stage at diagnosis
- First course of treatment
- Follow-up for vital status

VACCR Linkage with Other Person-Level Data is Possible...

- Linkage potential to utilization data
 - SSNs in VACCR Data
 - Link to the local tumor registries for SSNs
- Data use policies not well defined
- Research use of VACCR identifiable data is very limited

VACCR Data Quality

- Minimal empirically based evaluations
- GPRA Evaluation underway (FY2005)
- Ongoing validation study underway
 - Dominitz, et al., ERIC Funded

Questions?