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

June 20-24, 2005

# Mortality Data

# Session Objectives

- Identify mortality data sources for veterans available within the VA
- Understand the completeness and accuracy of VA mortality data
- Learn how multiple sources can be used to improve accuracy and completeness

# Session Objectives

- **Identify mortality data sources for veterans available within the VA**
- **Understand the completeness and accuracy of VA mortality data**
- **Learn how multiple sources can be used to improve accuracy and completeness**

# Mortality Data

- Five-year survival a common outcome measure in cancer research
- Short-term mortality is also used frequently
- Calculated using death date

# Data Sources Available in the VA

- VBA Beneficiary Identification and Record Locator Subsystem (BIRLS) Death File (DF)
- Medical SAS Inpatient Main Datasets (Patient Treatment Files or PTF)
- Social Security Administration (SSA) Death Master File (DMF)
- Medicare Vital Status File (MVS)

# BIRLS

- Veterans Benefits Administration (VBA) database
- Records of all VBA beneficiaries
  - Around 44 million records
  - Covers pre-Civil War to the present
  - Includes records of veterans whose survivors applied for death benefits
- Replaced the Master Index (a manual system) in 1972
- File on AAC Mainframe
- Includes date and cause of death
- VA collects death dates from
  - Survivors applying for death benefits
  - VHA hospitals
  - VA National Cemetery Administration

# BIRLS Death File (DF)

- VA's master record of death dates
- Is an extract of BIRLS
- Date and cause of death
- Cause of death sporadically populated
- Includes individuals who may have never been known to the VHA
- Approximately 12.7 m deaths as of April, 2005
- Both real and scrambled social security numbers
- Resides on AAC
- MDPPRD.PRO.SAS.VHA.BIRLS.DEATH

# Medical SAS Inpatient Datasets

- aka Patient Treatment Files or PTF
- Death dates from Oct 1991 – present
- In-hospital deaths & deaths reported to the hospital by a survivor
- Medical SAS Inpatient Datasets
  - Acute Main (90% of all PTF deaths)
  - Extended Main
  - Observation Main (9.5% of all PTF deaths)
  - Non-VA Main

# SSA Death Master File (DMF)

- Cumulative records of death dates reported to SSA
- Hill and Rosenwaike, *Social Security Bulletin* 2001/2002
  - Sensitivity varies by age groups (93 – 96% for 65yrs or older)
  - Sensitivity varies before and after 1973
- VA purchases DMF extract for internal use
- Contains over 70 million deaths
- Updated monthly

# SSA DMF

- Death reports from
  - family members
  - funeral homes
  - states and federal agencies
  - postal authorities
  - financial institutions

- MDPPRD.MDP.SSA.DEATH
  - Not a SAS Dataset
  - Layout available from the US Department of Commerce, National Technical Information Service (NTIS) https://dmf.ntis.gov/recordlayout.pdf

# Medicare Vital Status (MVS)

- Death dates for Medicare beneficiaries
- Cumulative file
- SSA DMF is the main source
- CMS collects death dates from Medicare claims data and other sources
- Source for other demographic data on Medicare-enrolled veterans
  - DOB
  - Sex, race, age,
  - State and county codes, zip code
  - Real and scrambled SSN

# MVS

- VIReC receives MVS annually
- VA researchers can request the file
- For more information about requesting data, go to the VIReC Web site: http://www.virec.research.med.va.gov

# Research Examples

- Rabeneck et al., *AJG* 2004
  - Outcomes of colorectal cancer surgery
  - 30-day and 5-year mortality
  - BIRLS-DF for death dates
  - Inpatient and outpatient datasets to identify cohort
  - 30-day mortality for rectal resection
    - 2.1% for < 65
    - 4.9% for ≥ 65
  - 5-year survival for rectal resection
    - 54.0% for < 65
    - 44.5% for ≥ 65
  - Mortality for those < 65 could be more severely underestimated than for those ≥ 65

# Research Examples

- Ho et al., *Ann Thorac Surg* 2005
  - Cardiac valve surgery patients
  - Compared 6-month mortality between depressed and non-depressed
  - BIRLS-DF for death dates
  - 13.2% unadjusted mortality for depressed vs. 7.6% non-depressed (Adjusted OR = 1.90; $p < 0.03$)

# Research Examples

- Fisher et al., *Am J Gastroenterol* 2003
  - Colorectal cancer patients
  - 5-year survival
  - BIRLS-DF was the source for death dates
  - 5-year survival rate was 69.1% for those with at least one follow-up colonoscopy, compared with 56.7% without any (Adjusted HR = 0.57; $p < 0.001$)

# Session Objectives

- Identify mortality data sources for veterans available within the VA
- **Understand the completeness and accuracy of VA mortality data**
- Learn how multiple sources can be used to improve accuracy and completeness

# Previous Studies Comparing Mortality Data

- Sensitivity of BIRLS-DF data
  - Highly variable among study cohorts (74 – 97%)
  - Fisher et al. *Am J Epidemiol* 1995 – 94.5% (cited in 20 published studies)
  - BIRLS-DF & PTF combined (Dominitz et al. 2001):
    - Inpatients – 91.1%
    - Outpatients - 73.5%
- Combined data improves sensitivity
  - BIRLS-DF & PTF (Dominitz et al. 2001)
  - BIRLS-DF & SSA DMF (Page et al. 1996)
- MVS has not been combined w/VA mortality data

# Feasibility Study

- HSR&D SDR 03-157
- Sohn MW, PI; Hynes DM; Arnold N; Maynard C
- January, 2004 – January, 2005
- Death dates from the National Death Index (NDI) too expensive
- Specific Aims:
  - To study costs and benefits of using alternative sources for death ascertainment
  - To develop a strategy for setting up a central registry for mortality data

# An NDI Search

- The study involves drawing small samples to search in the NDI database for comparison

- Establishing “match criteria” to determine the true match
  - SSN, Gender, 2 parts of DOB (Year, Month, or Day)
  - 3 other criteria

# National Death Index

- Database of state death records
- Widely recognized as the most reliable source of dates and causes of death
- Sensitivity over 95%
- Maintained by the National Center for Health Statistics
- Fee for NDI Search
- NDI *Plus* Search for cause of death

# Study Population

- All veterans known to the VA
  - VHA users since 1997
  - VHA enrollees
  - Compensation & Pension beneficiaries
- Excluded
  - Non-veterans
  - Records with invalid SSNs & no DOB
  - Deceased before January 1, 1999
  - Deceased after December 31, 2002
- 8.2 million veterans

# Samples Submitted to the NDI

- Simple random sample of 3,000 veterans
- Other samples for use to construct combined data were also submitted

| Description of Sample | Sample Size |
|---|---|
| Multiple sources with the same date of death | 2,376 |
| Multiple sources with differing dates of death | 2,767 |
| Only one source with a date of death | 2,154 |
| Presumed living | 2,000 |
| Unknown status | 4,000 |
| Activity after Death | 2,000 |

# Combined Mortality Data

- A file that combines death dates from
  - BIRLS Death File
  - PTF
  - SSA DMF
  - MVS
- One source with a date of death (7.5% of all deaths)
- Multiple sources with dates that agree (88.1%)
- Multiple sources with dates that do not agree (4.4%)
  - Computer routine was developed
  - Selects dates that maximize agreement with NDI dates
    - Sensitivity of the source
    - Any additional information
  - Routine reported in Technical Report

# Results

- Of 3,000 in the sample, NDI rejected 2
- Total sample size used in analysis = 2,998
- 292 deaths (9.7%) identified with NDI search

# Comparison with NDI Dates

| Source | Deaths | Sensitivity | Specificity | PPV* | NPV§ |
|---|---|---|---|---|---|
| BIRLS-DF | 226 | 76.4 | 99.9 | 98.7 | 97.5 |
| PTF | 35 | 12.0 | 100.0 | 100.0 | 91.3 |
| MVS¶ | 244 | 82.5 | 99.9 | 98.8 | 98.1 |
| DMF | 269 | 91.1 | 99.9 | 98.9 | 99.0 |
| Combined | 291 | 97.9 | 99.8 | 98.3 | 99.8 |

*PPV = positive predicted value; §NPV = negative predicted value.

¶Sensitivity for Medicare beneficiaries only was 98.4%.

# Sensitivity of Data Sources

- Poor sensitivity for BIRLS-DF – 76.4%
  - Mainly due to incomplete death reporting
  - 69 unreported deaths (23.6% of all deaths)
  - 3 deaths reported in BIRLS-DF, not in NDI
- Adequate sensitivity for SSA DMF – 91.1%
  - 26 unreported deaths (8.9% of all deaths)
  - 3 deaths reported in DMF, not in NDI
- High sensitivity for combined data – 97.9%
  - 6 unreported deaths (2.1% of all deaths)
  - 5 deaths reported in combined data, not in NDI
    - 2 dates in two or more sources and they agreed
    - 3 dates from only one source

# Sensitivity of Data Sources

- PTF
  - 12% for all veterans in the sample
  - 48% for all inpatient users
  - Need other sources of mortality data for inpatient users
- MVS
  - 82.5% for all veterans in the sample
  - 98.4% for Medicare beneficiaries in the sample
  - Comprehensive source for Medicare beneficiaries

# Agreement with NDI Dates

| Source | Number of Deaths | Death Dates in a Source that Agree with NDI Dates (%) | | |
|---|---|---|---|---|
| | | Exactly | Within 2 Days | In Year & Month |
| BIRLS-DF | 226 | 220 (97.3) | 222 (98.2) | 223 (98.7) |
| PTF | 35 | 35 (100.0) | 35 (100.0) | 35 (100.0) |
| MVS | 244 | 233 (95.5) | 237 (97.1) | 241 (98.8) |
| DMF | 269 | 257 (95.5) | 262 (97.4) | 266 (98.9) |
| Combined | 291 | 281 (96.6) | 284 (97.6) | 286 (98.3) |

# Accuracy

- Excellent accuracy for all sources

- Exact match with NDI dates
  - 97.3% of all death dates in BIRLS-DF
  - 100% for PTF
  - Over 95% for MVS and SSA DMF
  - 96.6% for the Combined Data

# Accuracy

- Match within 2 days
  - Over 97% of death dates in all sources agree with NDI dates
- Month and Year Agreement
  - Over 98% agreement in all sources
  - SSA DMF & MVS contain “plugged” dates when the actual day is not known
    - SSA DMF – 1st and 15th of the month
    - Medicare – end of month (28th, 30th, 31st)

# VHA Users vs. Non-Users

| User Groups* | N (%) | Sensitivity | | |
|---|---|---|---|---|
| | | BIRLS-DF | DMF | Combined |
| Inpatient Users | 342 (11.4) | 85.9 | 93.0 | 100.0 |
| Outpatient Users | 2,100 (70.0) | 79.1 | 92.5 | 100.0 |
| Non-Users | 894 (29.8) | 71.2 | 88.5 | 94.2 |

*Based on any use in CY1999 – CY2002

# Sensitivity by User Groups

- Sensitivity in BIRLS-DF
  - Highest (85.9%) for users of VA inpatient care
  - Lowest (71.2%) for non-users
- SSA DMF with less variations
  - 4.5% difference between highest and lowest groups
- Combined Data
  - Lowest with 94.2% sensitivity for non-users

# Age Groups

| Age Groups* | N (%) | Sensitivity | | |
|---|---|---|---|---|
| | | BIRLS-DF | DMF | Combined |
| < 65 | 1,762 (58.8) | 67.6 | 90.1 | 95.8 |
| ≥ 65 | 1,236 (41.2) | 79.2 | 91.4 | 98.6 |

*Age on January 1, 1999

# Sensitivity by Age Groups

- Sensitivity of mortality data in BIRLS-DF
  - Poor for both age groups
  - Higher for 65 or older
- SSA DMF
  - Respectable sensitivities
  - Little difference between the groups
- Combined data
  - Excellent sensitivities for both age groups

# Session Objectives

- Identify mortality data sources for veterans available within the VA
- Understand the completeness and accuracy of VA mortality data
- **Learn how multiple sources can be used to improve accuracy and completeness**

# Data Source for Combined Data

| | Source | | | |
|---|---|---|---|---|
| | BIRLS-DF | PTF | MVS | SSA |
| A. Death dates found in source | 226 | 35 | 244 | 269 |
| B. Death dates used in combined data | 221 | 35 | 240 | 264 |
| C. Unique dates found in source | 7 | 2 | 5 | 13 |
| D. % unique from source (C / A) | 3.2% | 5.7% | 2.1% | 4.9% |
| E. % unique of total deaths (C / 291) | 2.4% | 0.7% | 1.7% | 4.5% |

# Contribution to Combined Data

- All sources contributed to the combined data
- Combining four sources increased sensitivity rates compared with previous research
  - BIRLS-DF & PTF (Dominitz et al. 2001) – 91% for inpatients
  - BIRLS-DF & SSA DMF (Page et al. 1996) – 96%
- Surprisingly, SSA DMF is the source that contributed the largest number of death dates in the combined data

# In Summary

- BIRLS Death File
  - Poor in sensitivity but excellent in accuracy
  - Main problem is its incompleteness (23.6% of deaths not reported)
  - Large variations in sensitivity among cohorts
  - Variations in sensitivity by disease cohorts not known
  - Lower sensitivity rates for BIRLS-DF than previous studies due to non-users in sample
  - Needs other sources to supplement

- SSA DMF is the most complete single source of death dates for veterans (8.9% not reported)

- Carefully combined data are comparable to the NDI data in accuracy & completeness
  - Sensitivity = 97.9%
  - Accuracy = 96.6% exactly matched with NDI dates

- Researchers with access to these four sources do not need NDI searches for death dates

# Resources

- VIReC Web site – BIRLS-DF page
  - http://www.virec.research.med.va.gov/DataSourcesName/BIRLS/BIRLS.htm
- National Death Index
  - http://www.cdc.gov/nchs/r&d/ndi/ndi.htm
- SSA Epidemiological Search
  - Presumed living and unknown status
  - http://www.ssa.gov/policy/about/epidemiology.html

# Help on VA Mortality Data

- **VIReC Helpdesk**
  - **VIReC@va.gov**
  - **(708) 202 – 2413**

# Questions?