NCHS Record Linkage Program

National Conference on Health Statistics

August 16, 2010

NCHS Major Data Systems

- National Vital Statistics System and surveys
  - Linkage source for death certificate information through the National Death Index

- National Health Interview Survey (NHIS)

- National Health and Nutrition Examination Survey (NHANES)

- National Health Care Surveys

National Health Interview Survey

- Data source
  - Representative in person, in home survey of ≈ 87,500 participants
- Data applications
  - Health status and disability
  - Insurance coverage
  - Access to and use of health services
  - Extent of illness and disability
  - Immunization
  - Health behaviors
- Second Longitudinal Study of Aging (LSOA II)
  - Baseline data from 1994 NHIS

National Health and Nutrition Examination Survey

- Data source
  - Standardized physical examinations, laboratory tests, personal interviews with annual sample of 5,000
- Data applications
  - Disease or condition prevalence
  - Risk factors
  - Nutrition monitoring
  - Anthropometry
  - Growth and development
  - Disease monitoring

National Health Care Surveys

- Nursing Home Survey (NHHS)
  - Survey of nursing homes and their residents and has been in the field periodically since 1973
  - Provides information on nursing homes from two perspectives
    - The facilities
      - Size, ownership, Medicare/Medicaid certification, services provided and specialty programs offered, and charges
    - The recipients of care
      - Demographic characteristics, health status and medications taken, services received, and sources of payment.
  - 1985, 1995, 1997, and 2004 NNHS have been linked to various administrative data sources

NCHS Record Linkage Program

- Links health survey data with data collected from health related administrative records

- Designed to maximize the scientific value of the NCHS population-based surveys

- Examine factors that influence chronic disease, disability, health care utilization, morbidity, and mortality

NCHS Linkage Activities

- Mortality
  - National Death Index (NDI)
- Retirement and Disability
  - Social Security Administration (SSA)
- Medicare and Medicaid enrollment and claims
  - Centers for Medicare and Medicaid Services (CMS)
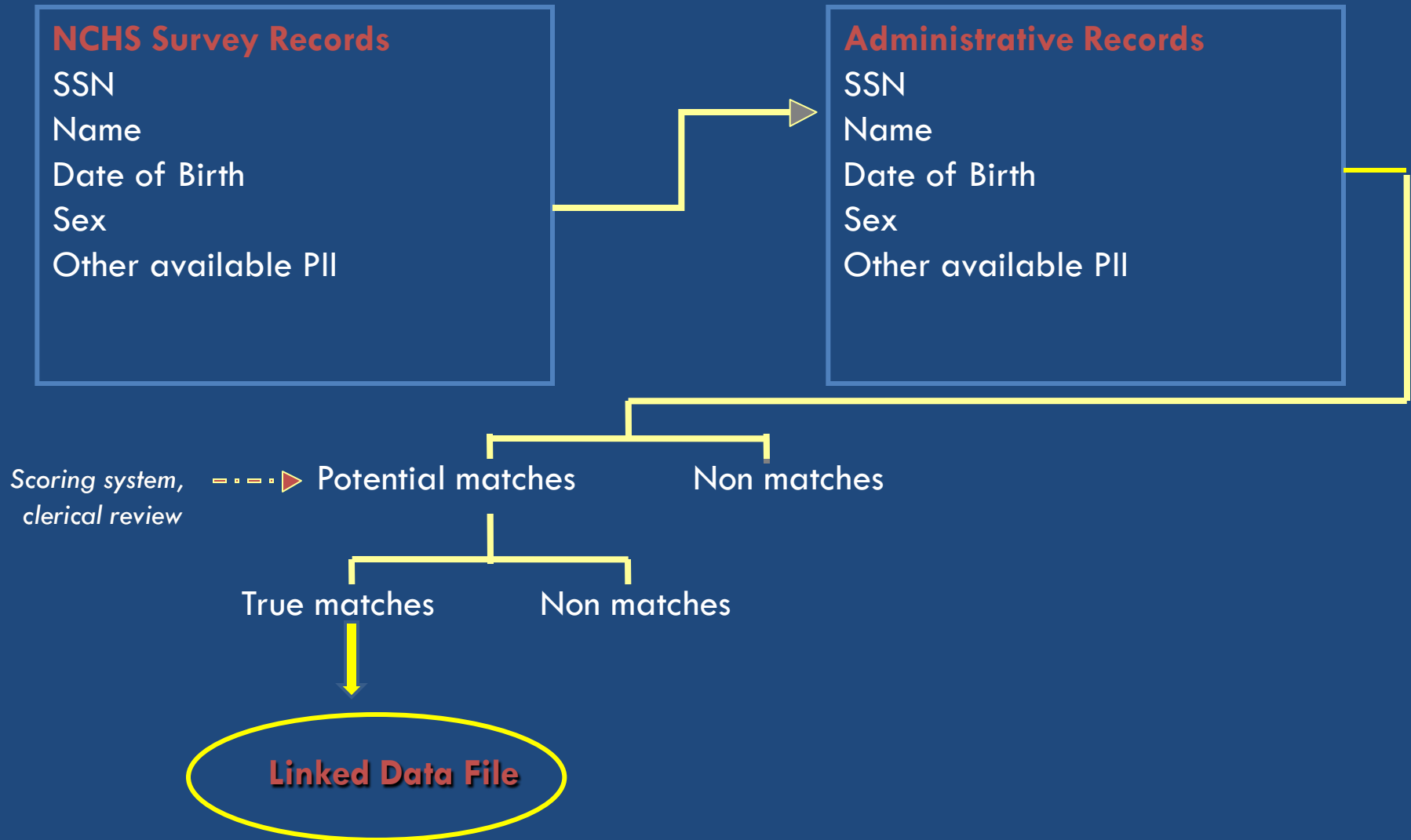- Air Quality
  - Environmental Protection Agency (EPA)

Why Do Linkage?

- Augments available information for major diseases, risk factors, and health service utilization
  - Links exposures to outcomes
  - Provides longitudinal component to survey data
- Reduces cost burden
  - Re-contacting survey respondents for follow-up information can be expensive
- Increases accuracy and detail of data collected

How Records are Linked

NCHS Survey Records
SSN
Name
Date of Birth
Sex
Other available PII

Administrative Records
SSN
Name
Date of Birth
Sex
Other available PII

Scoring system, clerical review ---▷ Potential matches   Non matches

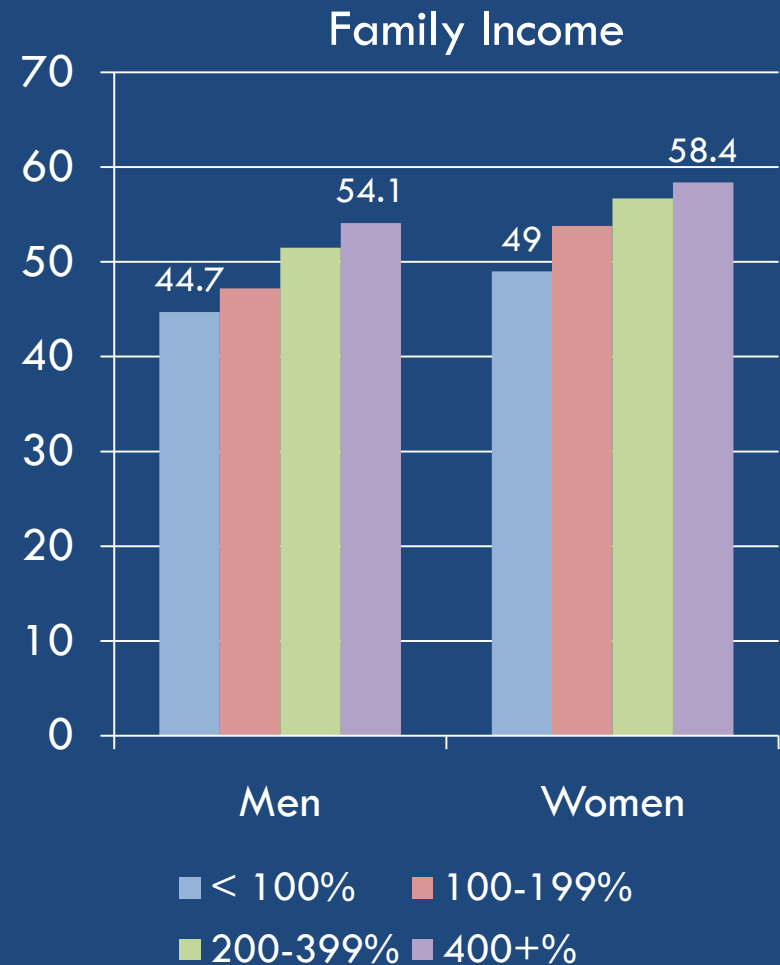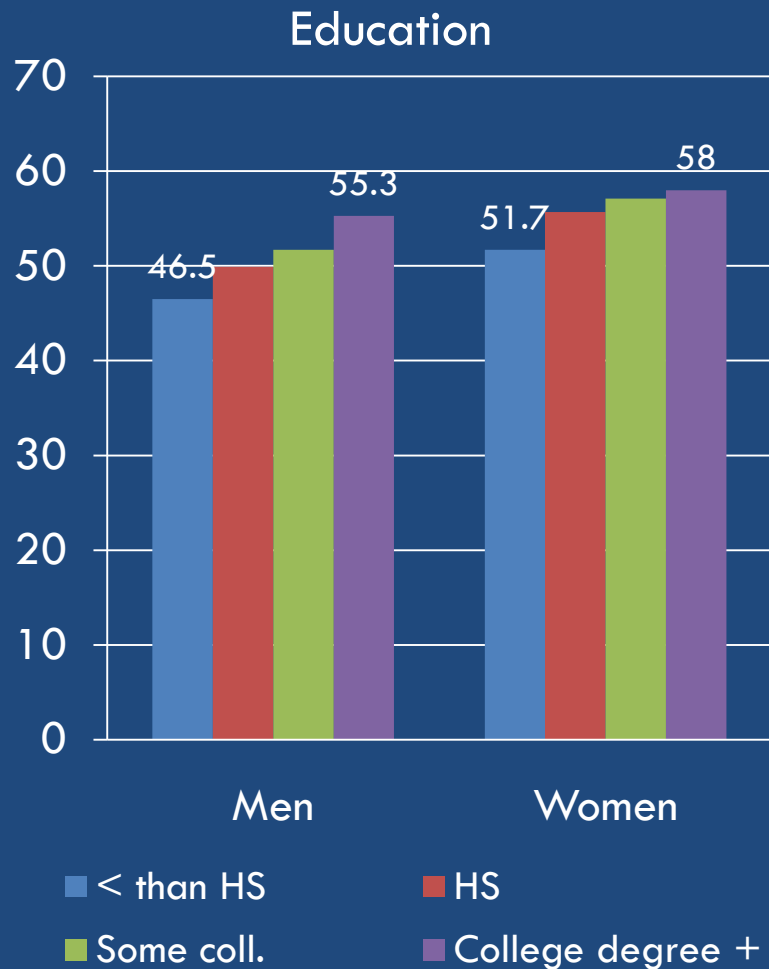True matches   Non matches

Linked Data File

Research Potential of NCHS Linked Data

- Disparities
  - Mortality patterns by race/ethnicity or socioeconomic status
- Health services
  - Impact of obesity on Medicare and Medicaid program expenditures
- Epidemiology
  - Impact of risk factors on health outcomes
- Disability
  - Effects of chronic illness and obesity on disability and mortality
- Methodologic studies
  - Validation of self-reports vs. administrative records
- Environmental Health
  - Effects of air pollution on health outcomes

Socioeconomic Differences in Life Expectancy at Age 25

Education

Men:
- < than HS: 46.5
- College degree +: 55.3

Women:
- < than HS: 51.7
- College degree +: 58

Legend:
- < than HS
- HS
- Some coll.
- College degree +

Family Income

Men:
- < 100%: 44.7
- 400+%: 54.1

Women:
- < 100%: 49
- 400+%: 58.4

Legend:
- < 100%
- 100-199%
- 200-399%
- 400+%

Source: 1990-2000 NHIS Linked Mortality Files, with mortality follow-up through 2002.

Longevity, BMI, and Lifetime Medicare Expenditures

BMI at age 45	Percent survived to age 65	LE at age 65 years	Estimated Lifetime Medicare Expenditures
Normal	86.3	19.5	$ 91,144
Over-weight	85.9	17.9	102,595
Obese	78.3	17.4	133,483

Year 2000 dollars, BMI is body mass index (Kg/m^2)
Data source: National Health and Nutrition Examination Survey Linked Medicare File

Effect of Vitamin D on Risk of Hip Fracture
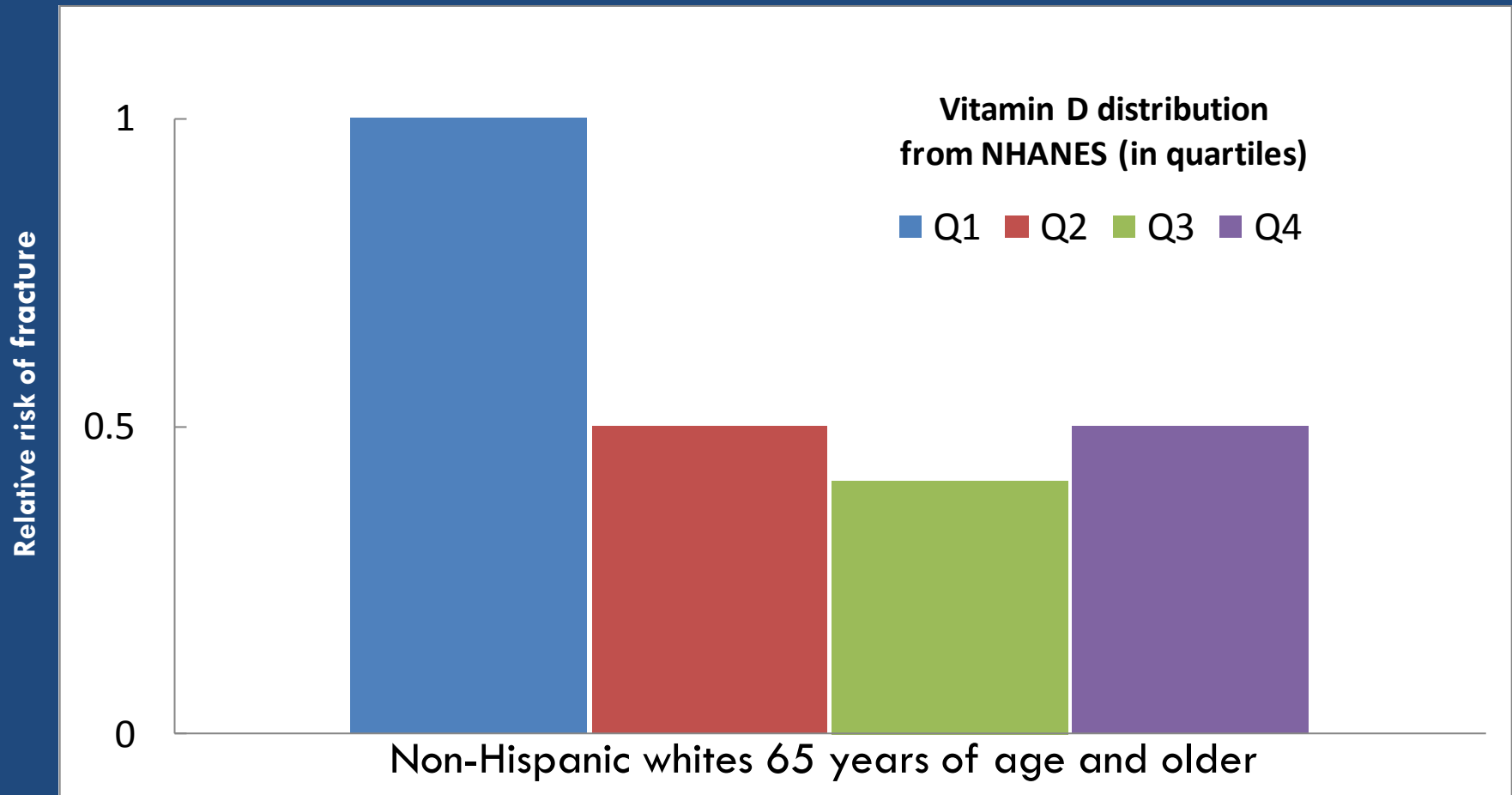
*Adjusted for age, sex, femur neck BMD, previous fractures, dietary calcium, kilocalories and weight loss from maximum
Source: NHANES III, Medicare Utilization and Expenditure Linked File; Looker et al J Bone Miner Res 2008;23:143.

Linked Mortality Files

NCHS Linked Mortality Files

NCHS Survey	Total Deaths through 2006
NHIS 1986-2004	181,188
LSOA II	5,806
NHEFS	8,002
NHANES II	5,276
NHANES III	5,478
NHANES 1999-2000	611
NHANES 2001-2002	470
NHANES 2003-2004	276
2004 NNHS	6,767
1997 NNHS	12,525
1995 NNHS	6,778
1985 NNHS	10,226

Linked Mortality Files:
Mortality Follow-up through 2006

	Restricted-use	Public-use
National Health Interview Survey (1986-2004)	X	X
Second Longitudinal Study of Aging (LSOA II)	X	X
NHANES I Epidemiologic Follow-up Study (NHEFS, 1971-1992)	X	
NHANES II (1976-1980)	X	
NHANES III (1988-1994)	X	X
NHANES (1999-2004)	X	X
1985, 1995, 1997 National Nursing Home Surveys (NNHS)	X	
2004 National Nursing Home Survey (NNHS)	X	X

Restricted-use vs. Public-use Files

☐ Restricted-use files

   ☐ Mortality follow-up for all eligible survey participants

   - Includes children for the NHIS, NHANES II, NHANES III, NHANES 1999-2004, and NNHS
   - Detailed date of death, birth date, and interview date
   - Underlying and multiple cause of death information
   - No perturbed information

Restricted-use vs. Public-use Files

- Public-use files
  - Mortality follow-up for all eligible adult survey participants
    - Date of death limited to quarter/year or a calculated follow-up variable
    - Underlying cause of death
      - MCOD information for diabetes, hypertension, and hip fracture
    - Perturbed information for date or cause of death for selected decedents
      - Vital status not perturbed
  - Comparative analyses published on NCHS website

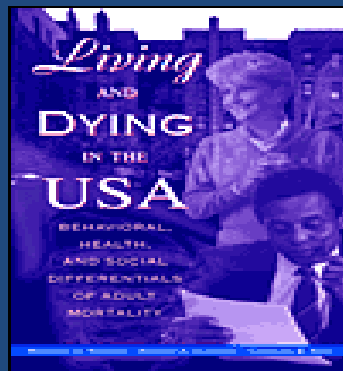# Research Potential of Linked Mortality Data

Excess Deaths Associated with Underweight, Overweight, and Obesity
KM Flegal, BI Graubard, DF Williamson, MH Gail;
JAMA, 2005;293:1861-1867.

Living and Dying in the USA:
Behavioral, Health, and Social
Differentials of Adult Mortality
RG Rogers, CB Nam, RA Hummer;
2000.

Suicide among male veterans: a prospective population-based study
MS Kaplan, N Huguet, BH McFarland, JT Newsom;
J Epidemiol Community Health, 2007; 61:619-624.

19

Linked Social Security Files

Linked Social Security Files

Social Security data spans 1953-2007

- Old Age, Survivors, and Disability Insurance (OASDI)
  - Master Beneficiary Record (MBR)
  - Payment History Update System (PHUS)
  - 831 Disability Master File (831 DMF)
  - Quarters of Coverage (Master Earnings File)
- Supplemental Security Income (SSI)
 - Supplemental Security Record (SSR)

Linked Social Security Files

- NCHS surveys
  - National Health Interview Survey (NHIS, 1994-2005)
  - Second Longitudinal Study of Aging (LSOA II, 1994-2000)
  - NHANES I Epidemiologic Follow-up Study (NHEFS, 1971-1992)
  - NHANES III (1988-1994)
  - NHANES (1999-2004)
  - National Nursing Home Survey (1985, 1995, 1997, 2004)

- Data access through the NCHS Research Data Center

Periods of Coverage for SSA Files
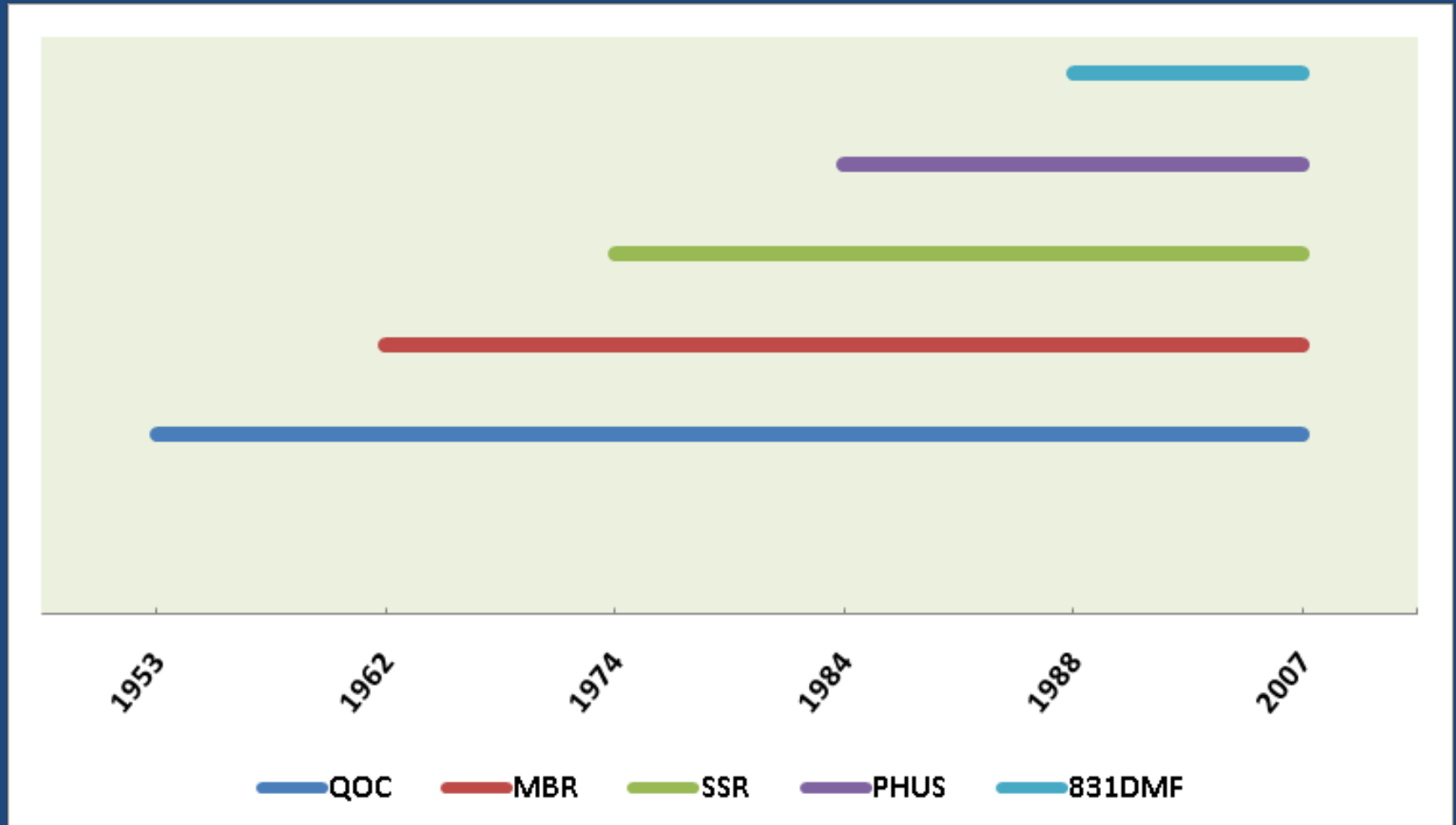
Social Security Data Elements

- MBR (1962-2007) and SSR (1974 -2007) files
  - Type of claimant
  - Benefit amount
  - Filing and eligibility status
  - Dual entitlement
- PHUS file (1984 -2007)
  - Actual monthly amount that a beneficiary received
  - Amount of monthly Medicare premium

Social Security Data Elements

- 831 Disability Master file (1989-2007)
  - Status of disability application, e.g. filing date
  - Diagnosis coding, e.g. primary and secondary impairments
- Quarters of Coverage (1953-2007)
  - Summarized variables indicating an individual's insured status, i.e. eligibility for receiving retirement or disability benefits

Research Potential of Linked Social Security Data

- Examine reliability of survey information for SSA program participation and benefits

- Compare the health characteristics of early retirees (age 62) to those who postpone benefits

- Policy analysis using validated survey data
  - Predicting the number of people who will become disabled based upon survey reported health conditions
  - Determining whether current disability entitlement funding levels will be adequate in the future

Linked CMS Files

Linked Medicare and Medicaid Files

CMS data spans 1991-2007

THESE UPDATED FILES EXPECTED BY END 2010

- Medicare Standard Analytic Files (SAF)
  - Denominator file
  - MEDPAR
  - Hospital outpatient
  - Home Health Agency
  - Hospice
  - Carrier (physician/supplier Part B file)
  - Durable Medical Equipment
- Summary Medicare Enrollment and Claims Files (SMEC)

Linked Medicare and Medicaid Files

THESE UPDATED FILES EXPECTED BY MID 2011

- Medicare Part D Prescription Drug Data
- Chronic Condition Warehouse (CCW) summary files
- End Stage Renal Disease (ESRD)
- Medicaid Analytic eXtract (MAX)

Linked Medicare and Medicaid Files

- NCHS surveys
  - National Health Interview Survey (NHIS,1994-2005)
  - Second Longitudinal Study of Aging (LSOA II, 1994-2000)
  - NHANES I Epidemiologic Follow-up Study (NHEFS, 1971-1992)
  - NHANES II (1976-1980)
  - NHANES III (1988-1994)
  - NHANES (1999-2004)
  - 2004 National Nursing Home Survey
- Data access through the NCHS Research Data Center
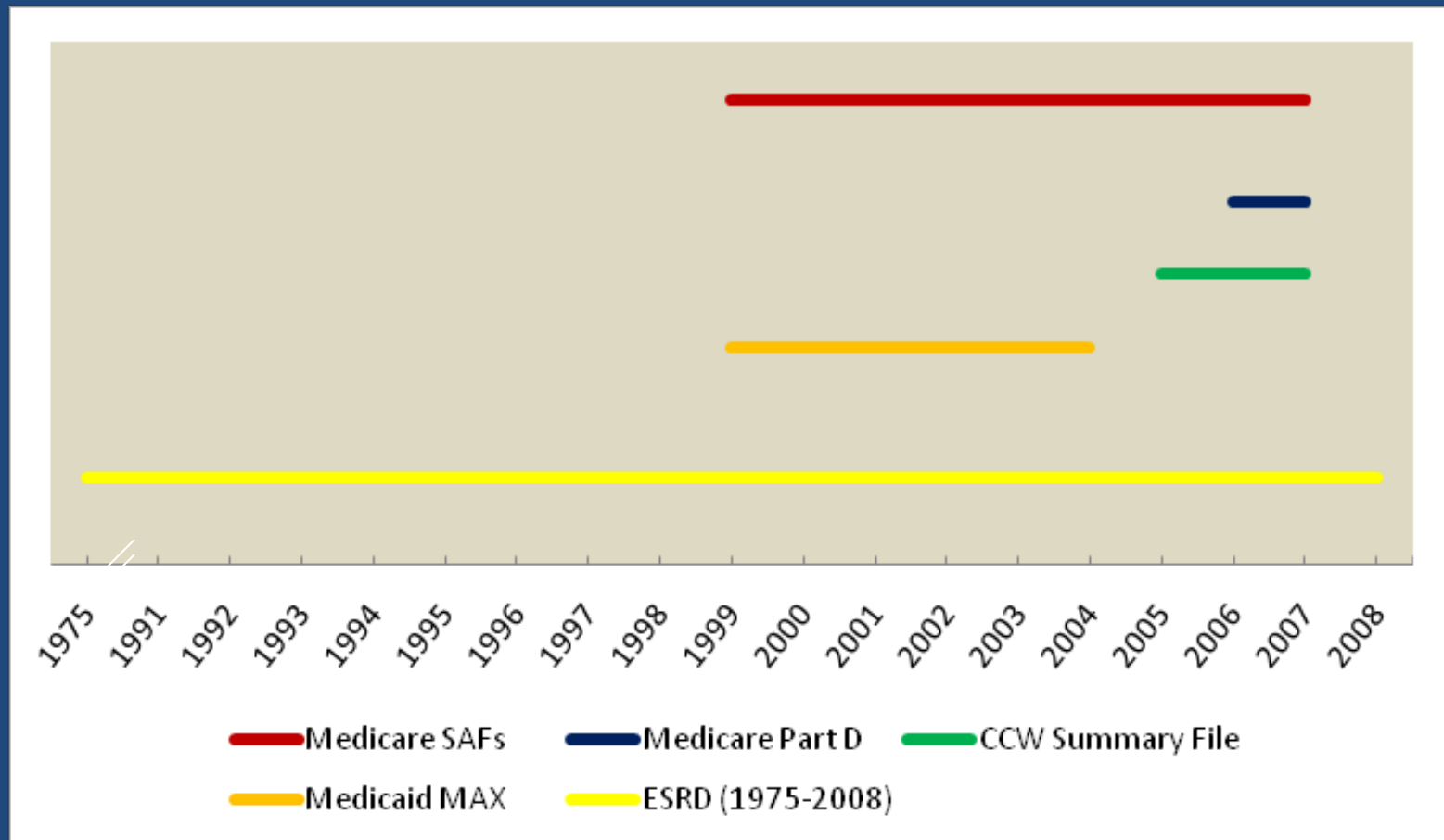
Periods of Coverage for CMS Files

Surveys: 1994-1998 NHIS, NHEFS, NHANES II, NHANES III, LSOA II
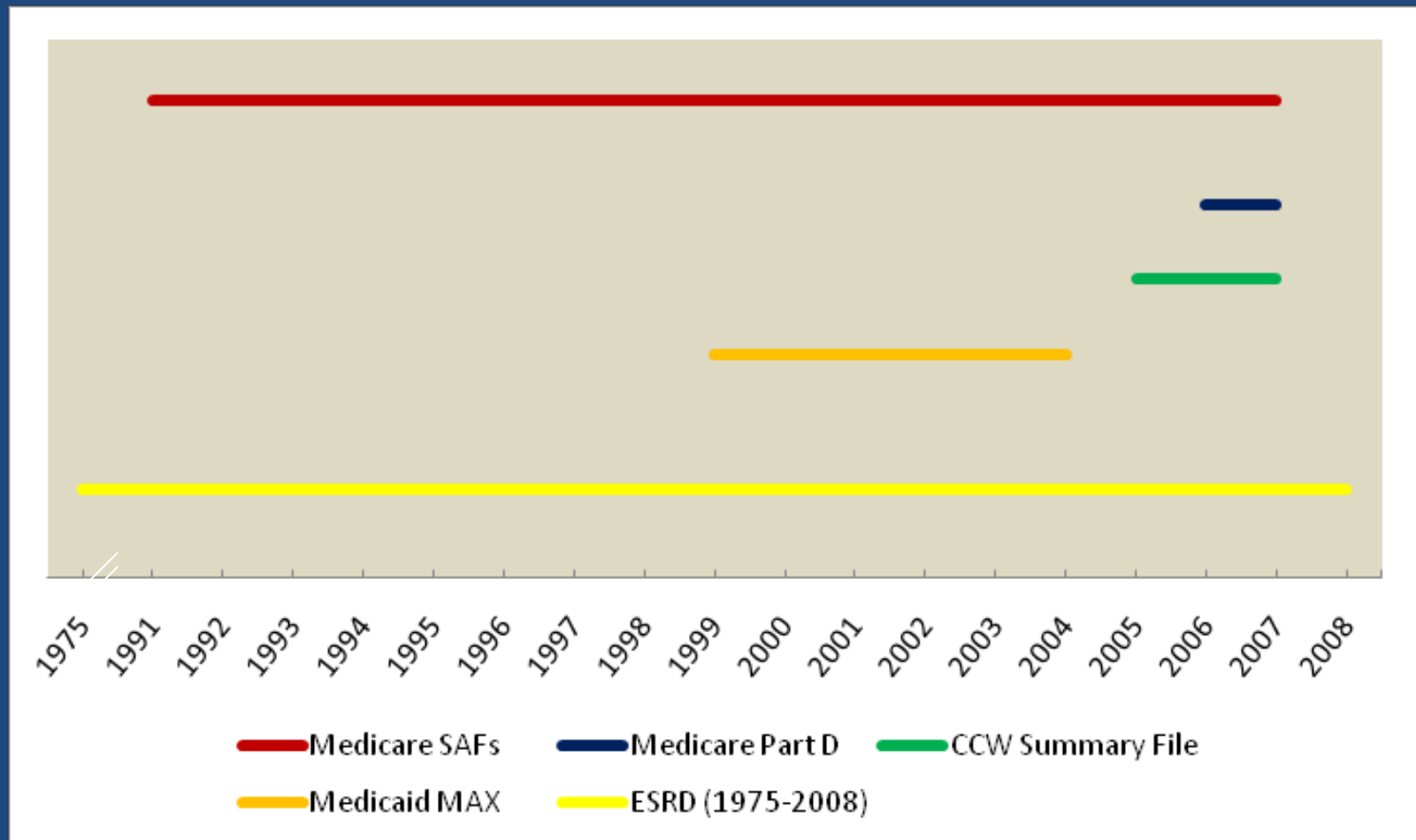
Periods of Coverage for CMS Files

Surveys: 1999-2005 NHIS, 1999-2004 NHANES, 2004 NNHS

Medicare Data Elements

- Denominator file (1991-2007)
  - Monthly enrollment status
  - Monthly HMO enrollment
  - CMS date of death
- Standard analytic files (1991-2007)
  - Diagnoses codes
  - Service dates
  - Reimbursement amount
  - Charge amount

Medicare Data Elements

- Medicare Part D (2006-2007)
  - Denominator file
    - Demographics
    - Enrollment
    - Subsidy information, e.g. low income, retiree
  - Drug event file
    - Summary extracts, no individual claims
    - Drug costs and payments submitted by Part D plan providers
    - Quantity dispensed and days supply
    - Patient paid amount

Medicare Data Elements

- CCW summary file (2005-2007)
  - Clinical information for 21 chronic health conditions
    - Acute MI, Alzheimer's, Alzheimer's related/senile dementia, atrial fibrillation, cataract, chronic kidney disease, COPD, depression, diabetes, glaucoma, heart failure, hip/pelvic fracture, ischemic heart disease, osteoporosis, RA/OA, Stroke, breast cancer, colorectal cancer, prostate cancer, lung cancer, endometrial cancer
  - Chronic condition flags
    - Yearly (Dec. 31 end of the reference year)
    - Mid-year (July 1 end of the reference year)
    - Ever (first occurrence of condition)
  - Yearly and mid-year flags provide indication of whether claims and/or coverage were met

Medicare Data Elements

- ESRD (1975-2008)
  - Combined ESRD Patient Master and Death Notification File
    - Primary and contributing causes of ESRD diagnosis
    - Date of first dialysis and information on kidney transplants
  - ESRD Entitlement/Registration file
    - Information from Medicare Entitlement and/or Registration (Form 2728)
    - Health insurance status
    - Primary and contributing causes of ESRD diagnosis
    - ESRD-related test results, dialysis information, and kidney transplant status
  - ESRD Patient Event file
    - ESRD related events, e.g. event date, type of event, and patient modality

Summary Medicare Data File

- Summary Medicare Enrollment and Claims file (SMEC) for 1991-2007
  - Enrollment information from the Denominator file plus summary variables of claims and payments
  - Variables modeled after MCBS cost and use files
    - Total reimbursements per year
    - Total number of claims by Medicare record type
    - Summary of charges by Medicare record type
    - Termination status & reason for termination
    - Monthly HMO enrollment
    - Medicare status code (i.e. Part A, B or both)

Medicaid Data Elements

- Medicaid (MAX, 1999-2004)
  - Medicaid eligibility, service utilization and payment
- Person Summary
  - Eligibility, demographics, HMO enrollment, summary of utilization and Medicaid payment by type of service
- Inpatient Hospitalization
  - Complete hospital stay records, e.g. diagnoses, procedures, discharge status, length of stay and payment
- Long Term Care
  - Claims records that include facility type, dates of service and discharge status
- Other Therapy
  - Claims for all non-institutional Medicaid services, e.g. physician services, lab/x-ray and premium payments
- Prescription Drug
  - Final action claims for paid prescription drugs
  - One procedure code but no diagnoses codes

Research Potential of Linked Medicare and Medicaid Data

- ☐ Examine risk factors for health conditions

- ☐ Examine uninsured estimates

- ☐ Examine disparities in Medicare service utilization

- ☐ Examine effect of obesity or Medicare and Medicaid costs

Linked EPA Air Quality Data Files

- Annual pollution averages collected from air monitoring locations
- Linked by geographic codes (e.g. county or zip code)
- NCHS surveys
  - National Health Interview Survey (1986-2005)
  - NHANES III (1986-1994)
  - National Hospital Discharge Survey (1999-2005)

Analytic Issues

Analytic Issues (selected)

- General
  - Survey respondents ineligible for linkage
  - Variance estimation
  - Issues with administrative data

- Mortality
  - Changes in data over time
  - Public-use files

Analytic Issues (selected)

- Medicare
  - Death
  - Linked but no Medicare data
  - Managed care enrollment
  - Study feasibility

- Social Security
  - Linked but no benefit history data
  - Study feasibility

Ineligible Respondents

- Cannot attempt to link survey records to administrative data sources if respondent
  - Refused to provide SSN (SSA or CMS only)
  - Lacks sufficient key identifying information

- These cases are INELIGIBLE for linkage and MUST BE DROPPED from all analysis
  - They are not a random sample of respondents

Ineligible Respondents

- Linked Mortality Files
  - Identify using ELIGSTAT
  - Ineligibility a problem for NHIS
    - Created new sample weights to account for ineligibility due to insufficient identifying data
- Linked CMS and SSA Files
  - Increasing ineligibility among NHIS
    - But match rate among those eligible remains high
  - No eligibility adjusted sample weights provided
  - Identify using the variable (CMS_MATCH or SSA_MATCH) on the Feasibility Study Data

Variance Estimation

- In general, follow guidelines and recommendations of the specific survey
- Pooled analyses of NHIS Files
  - Can span different NHIS design periods
  - Create new (consistent) design variables
    - Different design periods have different variable names and stratum identifiers have different lengths
  - Follow instructions available at
    - www.cdc.gov/nchs/data/nhis/9705var.pdf

Issues with SSA and CMS Administrative Data

- Created for non-research purposes (i.e. program administration)
  - Undocumented variation (e.g. among reporting entities)
  - Administrative data updates
  - Lack of documentation on data editing and/or imputation procedures
  - Very large files which may include many non-relevant variables

Linked Mortality Files:
Changes in Data Over Time

- Two sets of ICD codes cover linked mortality files
  - ICD-9 (deaths 1979 – 1998)
  - ICD-10 (deaths 1999 to present)
- Use both sets of codes to obtain full counts of cause-specific deaths
  - Restricted-use data includes individual codes
- Recoded deaths before 1998 using ICD-10 guidelines
  - UCOD_113 (public and restricted use files)
- Refer to vital statistics report on ICD comparability

Linked Mortality Files: Public-use data

- Public-use files analytically suitable for many types of analyses
  - Vital status not perturbed
  - Date or cause of death may be perturbed
- Comparative methods
  - Compared mean follow-up times and distributions for select causes of death for the public-use and restricted-use data
  - Compared the mortality risk for a standard set of socio-demographic covariates for all-cause as well as cause-specific mortality

Linked Mortality Files: Public-use data

- Public-use files yield similar hazard ratios and confidence intervals, particularly for common causes of death

- Results for less common causes of death remain consistent, although less agreement in the estimates

- Caution is urged for analyses of very rare causes of death or small population subgroups

- Users of the public-use linked mortality files may request to verify their results through the NCHS Research Data Center

Linked Medicare Files: Identifying Deaths

- Survey participants interviewed before the availability of linked Medicare files could have died (e.g. NHEFS, NHANES II or NHANES III)

  - Survey respondents who died before linked Medicare data can be identified by merging mortality information from the Linked Mortality files

  - Needed to create analytic sample

- Persons may die during study period and cease to have Medicare records

  - Look for a CMS date of death on each of the Denominator or SMEC files

Linked Medicare Files:
No Denominator Record

☐ Why does not having a denominator record matter?

☐ Lack of denominator record can affect your analytic sample – why?

  ◻ Can't determine managed care enrollment

  ◻ In general, managed care enrollees are excluded from sample (more on this to come)

Linked Medicare Files:
No Denominator Record

- Deceased
  - Prior to availability of linked Medicare data
    - Check mortality through linked mortality files
  - During periods of coverage for linked files
    - Check CMS date of death
- Loss of entitlement
  - check termination codes on Denominator or SMEC files (PART_A_TERM_CD and PART_B_TERM_CD)
    - Deceased
    - Non payment of premium
    - Voluntary withdrawal
    - Other termination
- CMS record keeping inconsistencies

Linked Medicare Files:
Denominator record but no claims data

- Entitled to Medicare, but not utilizing Medicare services during period that linked files cover
- Entitled to Medicare, but services being used are not reimbursable claims
- Note:
  - Small fraction of linked participants may have claims data but no associated Denominator record
  - Some Medicare beneficiaries only enrolled in Part A
- CMS record keeping inconsistencies

Linked Medicare Files:
Managed Care Enrollment

- No claims submitted for beneficiaries enrolled in managed care plans (HMO)
  - Do not have complete information on payments or services received
  - Could miss health events that are being counted based upon submitted claims
- Complex issue. See ResDAC
  - Technical brief at
  - www.resdac.umn.edu/Tools/TBs/TN-009.asp
  - Other studies www.resdac.umn.edu/AboutUs/Staff.asp

Linked Medicare Files: Managed Care Enrollment

- How managed care enrollees affect your research depends upon your question…
  - Studies on reimbursements/charges
    - Option may be to exclude those with any managed care enrollment because you don't have complete information on payments or services received
  - Studies on health outcomes/events
    - Option may be to exclude those with any managed care enrollment because you could miss events
    - Option may be to censor observations at time of first HMO enrollment
  - Other methods for addressing HMO enrollment possible depending upon research question

CMS Feasibility Study Data

- Public-use data
    - Indicates whether respondent was linked to CMS data (CMS_MATCH)
    - Provides limited variables on whether linked respondent has data on any of administrative files
        - Determine maximum sample size for each linked file
- Does not contain specific information, e.g. no information about amount or types of claims or benefits
- Access data from data linkage website

Linked but no SSA Data

- Linkage is to SSA NUMIDENT file
- Linked to NUMIDENT file but may not be eligible for Social Security benefits
  - Not age eligible for retirement
  - Defer retirement benefits because working full-time
  - Not eligible for Social Security

Issues with SSA Administrative Data

- Administrative data updates
  - Payment history updates
  - Previously denied claims may be overridden
  - Changes to type of benefit status
    - Individuals receiving disability (DI) switch to retirement (R) benefits at age 65 in RSDI program
- Complicated data
  - File layouts are complex, e.g. each MBR record has 2 parts
  - Calculation of benefits not straightforward, e.g. SSI benefits come from both federal and state programs

SSA Feasibility Study Data

- Public-use data
  - Indicates whether respondent was linked to SSA data (SSA_MATCH)
  - Provides limited variables on whether linked respondent has data on any of administrative files
    - Determine maximum sample size for each linked file
- Does not contain specific information, e.g. no information about amount or types of benefits
- Access data from data linkage website

Data User Tools

- File layouts & detailed notes
- Sample SAS & STATA input statements for public-use linked mortality files
- Matching methodology reports
- Linkage rates for SSA & CMS linked data
- Analytic guidelines
- Feasibility data files for SSA & CMS Files - Download from web
- Comparative analysis of the public-use and restricted-use linked mortality data

Additional Information

- Contact us at: datalinkage@cdc.gov
- NCHS data linkage activities
  - www.cdc.gov/nchs/data_access/data_linkage_activities.htm
- Mortality linkage
  - www.cdc.gov/nchs/data_access/data_linkage/mortality.htm
- SSA linkage
  - www.cdc.gov/nchs/data_access/data_linkage/ssa.htm
- CMS linkage
  - www.cdc.gov/nchs/data_access/data_linkage/cms.htm
  - Contact ResDAC - www.resdac.umn.edu/

Data Access - Data Linkage Activities - Microsoft Internet Explorer

File   Edit   View   Favorites   Tools   Help

Back

Address   http://www.cdc.gov/nchs/data_access/data_linkage_activities.htm   Go   Links

Google   Search   Share   Sidewiki   Bookmarks   Check   Translate   AutoFill   Sign In

Norton   Norton Safe Search   Search   Cards & Log-ins   LastPas

Data Access

National Center for Health Statistics

Data Access

Data Access Tools

▶ **Data Linkage**

Air Quality

Mortality Data

Medicare Enrollment and Claims Data

Social Security Benefit History Data

National Death Index

Public-Use Data Files

Research Data Center

Important Information

Data Release Policy

Data User Agreement

NCHS Home > Data Access

NCHS Data Linkage Activities

NCHS has developed a record linkage program designed to maximize the scientific value of the Center's population-based surveys. Linked data files enable researchers to examine the factors that influence disability, chronic disease, health care utilization, morbidity, and mortality. NCHS is currently linking various NCHS surveys with air monitoring data from the Environmental Protection Agency (EPA), death certificate records from the National Death Index (NDI), Medicare enrollment and claims data from the Centers for Medicare and Medicaid Services (CMS), and Retirement, Survivor, and Disability Insurance (RSDI) and Supplemental Security Income (SSI) benefit data from the Social Security Administration (SSA).

Linked Data Resources by Topic Area:

- Air Quality
- Mortality data
- Medicare Enrollment and Claims data (CMS)
- Social Security Benefit History data (SSA)

Email   Print   Share   PDF Reader   PPT Reader

Page last updated: March 11, 2009
Page last reviewed: February 24, 2009
Content source: CDC/National Center for Health Statistics
Page maintained by: Office of Information Services

Text size: S M L XL

Email page
Print page
Bookmark and share

Contact Us:

National Center for Health Statistics
3311 Toledo Rd
Hyattsville, MD 20782
1 (800) 232-4636
cdcinfo@cdc.gov

Data Access

Session 24. The Research Data Center (RDC): Accessing Restricted Variables From NCHS

TODAY: 4:00 p.m. – 5:30 p.m.

Data Linkage Restricted Files Include…

☐ Linked administrative data

   ☐ Medicare and Medicaid

   ☐ Social Security benefits

   ☐ Restricted-use linked mortality files

☐ EPA air pollution data

Why can't you just give me the data?

- NCHS does not "own" the linked administrative data

- NCHS data confidentiality rules prohibit the release of potentially identifiable data – special considerations concerning the protection of linked data

- The RDC is the only option for access to restricted-use data files

Data Access: The Research Data Center

- Identify a need for restricted data
- Submit a proposal
- Choose a mode of access
- Assigned an RDC Analyst
- Review Committee
- Compile public data
- Complete confidentiality training and paperwork
- Fees

Mode of Access

- NCHS RDC
  - Hyattsville, MD
  - Atlanta, GA (CDC)
- Census RDC – 10+ locations
- Remote Access System (ANDRE)
- Staff Assisted
- Combination of any of the above

- RDC provides access to the data
- RDC never gives you the dataset

Compile public data

- Researcher provides public data files
    - E.g. Demographic and covariate information from public-use survey files or some other non NCHS data source
    - Merged into NCHS restricted data with survey unique ID
- Researcher supplied data dictionary required with proposal
    - Proc contents = data dictionary
- Cannot include all variables – only those needed to answer this specific research question
    - **Cannot include public-use mortality variables with restricted-use mortality data**

Confidentiality and Disclosure Review

- Data never leave the secure access points

- Released output must be
  - reviewed first by RDC Analyst/ANDRE
  - able to answer the question in your proposal
  - final output intended for publication

- Data, Programs, and all Output remain on the computer for 1-2 years

RDC Resources

- Peter Meyer, MPH, MA
  - Director, Research Data Center
  - 301-458-4375
  - pmeyer1@cdc.gov

- Stephanie Robinson, MPH
  - Health Research Analyst
  - 770-488-5092
  - srobinson7@cdc.gov