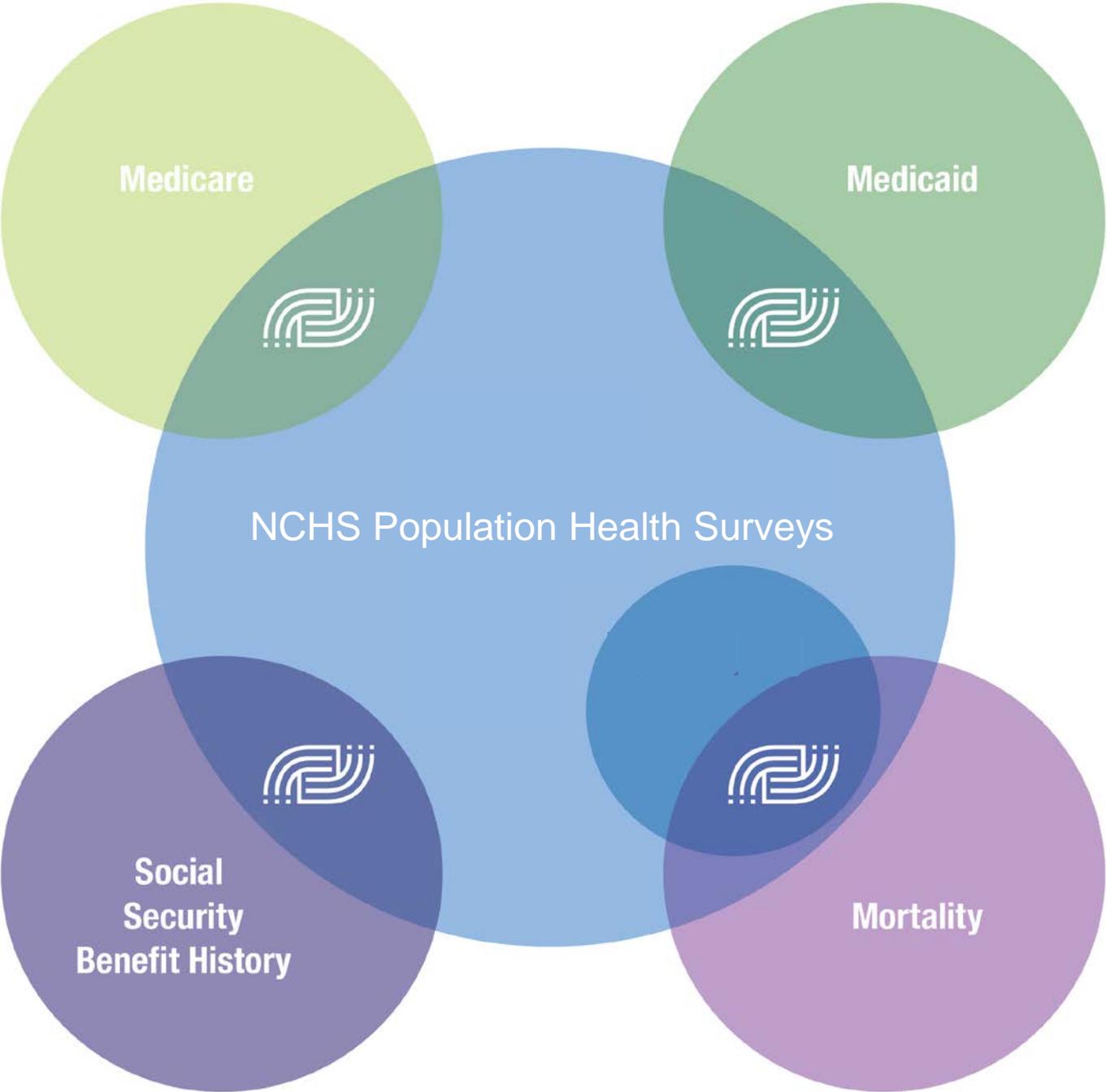# Overview of NCHS Linked Data and Analytic Considerations

Hannah R. Day, PhD

Objectives

- Overview of linked data resources
- Some analytic considerations of the linked data
- Motivating Example
  - Compare self-report diabetes in NHIS 2005 with 2005 Medicare claims for diabetes in the Chronic Condition (CC) Summary File.
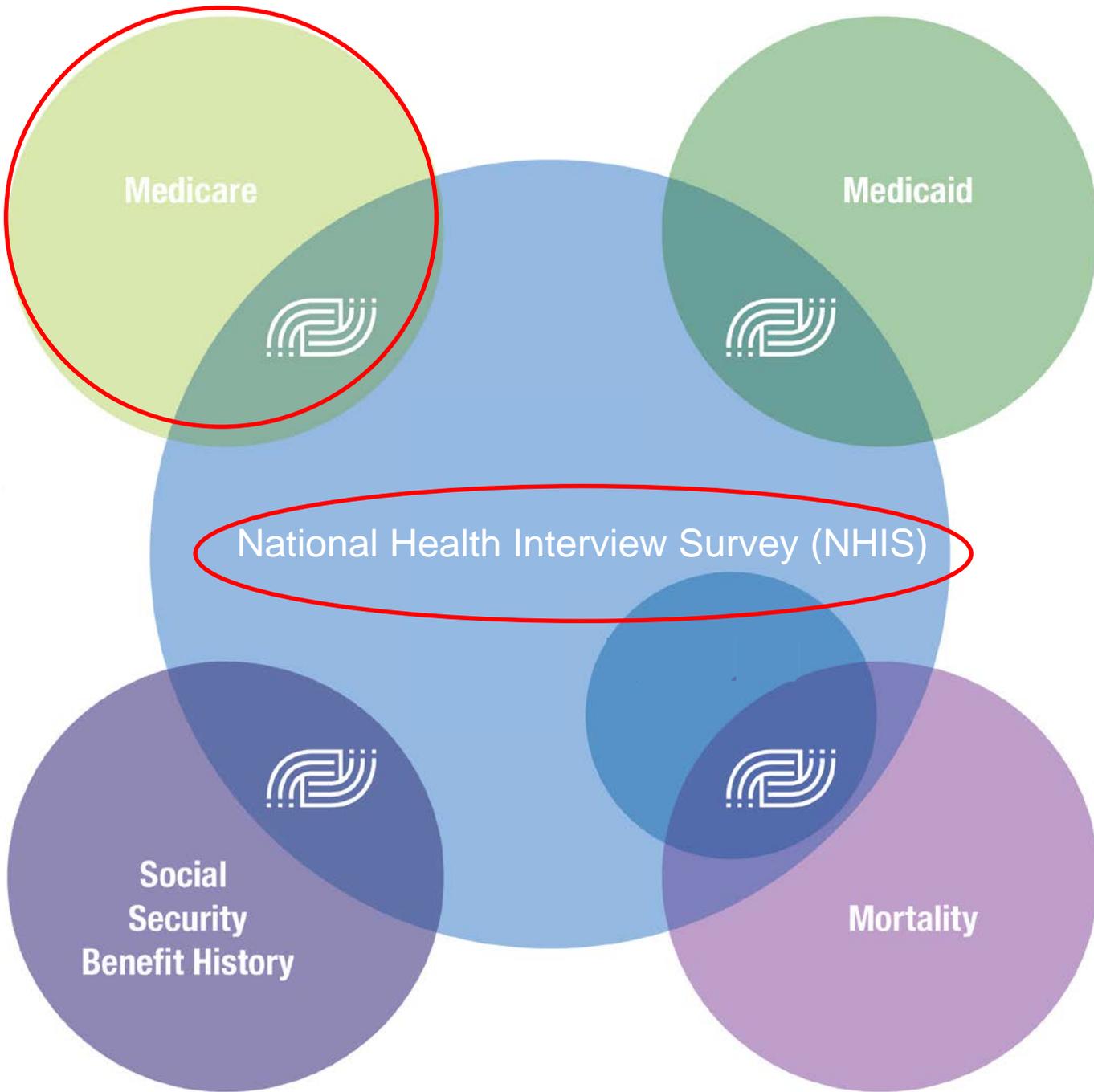
Surveys with Linked Administrative Data

	Survey
NHIS	National Health Interview Survey
NHANES	Continuous National Health and Nutrition Examination Survey
NHANES III	Third National Health and Nutrition Examination Survey
NHANES II	Second National Health and Nutrition Examination Survey
NHEFS	NHANES I Epidemiologic Followup Study
LSOA II	Second Longitudinal Studies of Aging
NNHS	National Nursing Home Surveys

Linked Data Considerations

- Defining Study Sample
  - Timing of survey data and administrative claims
  - Mortality
  - Program ineligible
- Data Issues
  - *Ever* Variable in Chronic Condition (CC) Summary File
  - Medicare Part C
- Incomplete Linkage
  - Linkage ineligible
  - Reweighting
- Child Survey Participants

Linked Data Considerations

- Defining Study Sample
  - Timing of survey data and administrative claims
  - Mortality
  - Program ineligible
- Data Issues
  - *Ever* Variable in Chronic Condition (CC) Summary File
  - Medicare Part C
- Incomplete Linkage
  - Linkage ineligible
  - Reweighting
- Child Survey Participants

Linked Data Considerations

- Defining Study Sample
  - Timing of survey data and administrative claims
  - Mortality
  - Program ineligible
- Data Issues
  - *Ever* Variable in Chronic Condition (CC) Summary File
  - Medicare Part C
- Incomplete Linkage
  - Linkage ineligible
  - Reweighting
- Child Survey Participants

Years of NHIS-Medicare linkage

Years of NHIS linked to Medicare:

1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
(2005)

Years of Medicare Data:

1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
Coming Soon:
2008
2009

9

Years of NHIS-Medicare linkage

Years of NHIS linked to Medicare:

1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005

Years of Medicare Data:

1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007

Coming Soon:
2008
2009

10

Linked Data Considerations

- Defining Study Sample
  - Timing of survey data and administrative claims
  - Mortality
  - Program ineligible
- Data Issues
  - *Ever* Variable in Chronic Condition (CC) Summary File
  - Medicare Part C
- Incomplete Linkage
  - Linkage ineligible
  - Reweighting
- Child Survey Participants

Years of NHIS-Medicare linkage

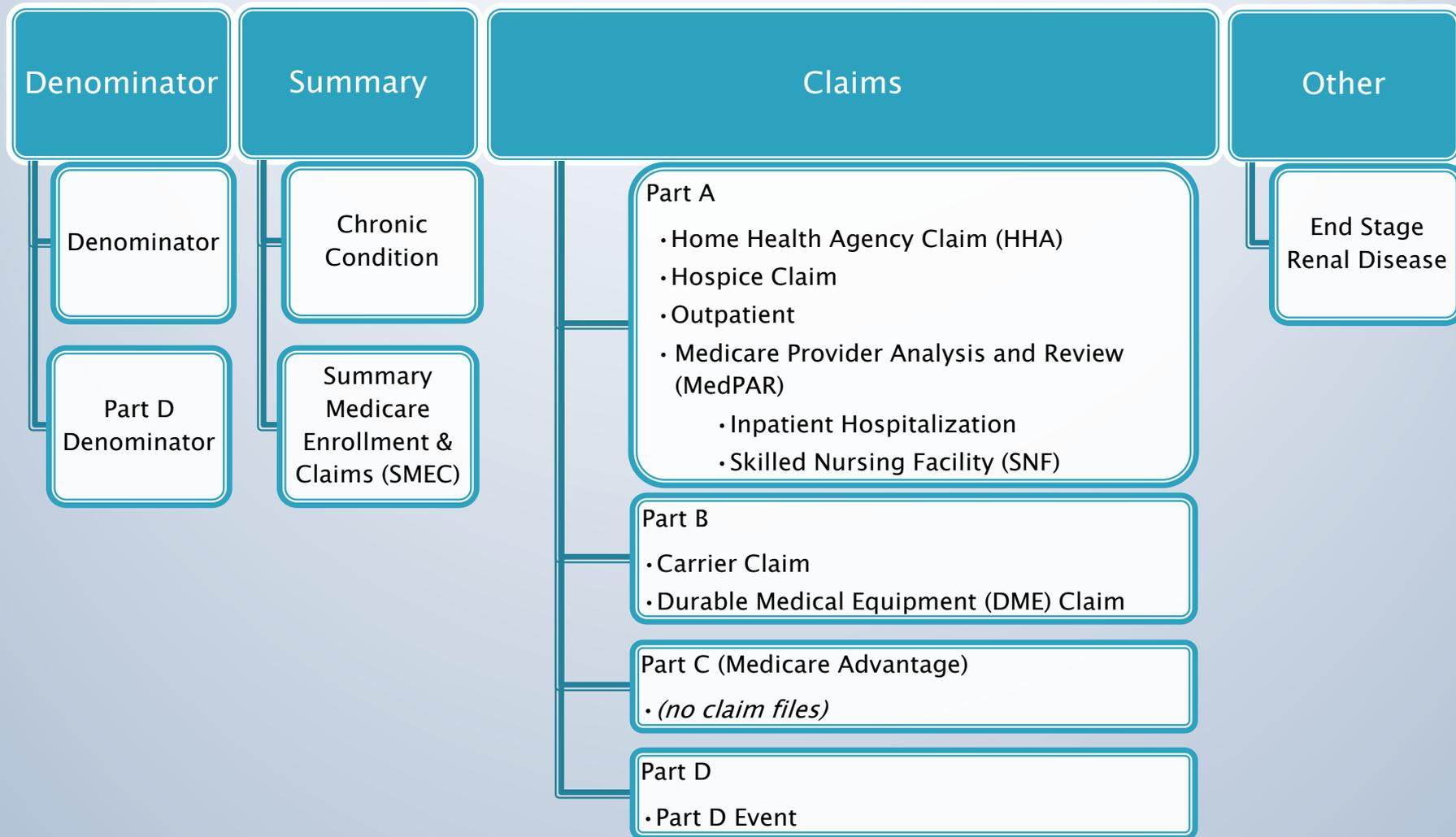| Years of NHIS linked to Medicare:	Years of NHIS linked to Mortality through 2006	Years of Medicare Data:
		1991
		1992
		1993
1994	1994	1994
1995	1995	1995
1996	1996	1996
1997	1997	1997
1998	1998	1998
1999	1999	1999
2000	2000	2000
2001	2001	2001
2002	2002	2002
2003	2003	2003
2004	2004	2004
2005		2005
		2006
		2007

Coming Soon:
2008
2009

12

Linked Data Considerations

- Defining Study Sample
  - Timing of survey data and administrative claims
  - Mortality
  - Program ineligible
- Data Issues
  - *Ever* Variable in Chronic Condition (CC) Summary File
  - Medicare Part C
- Incomplete Linkage
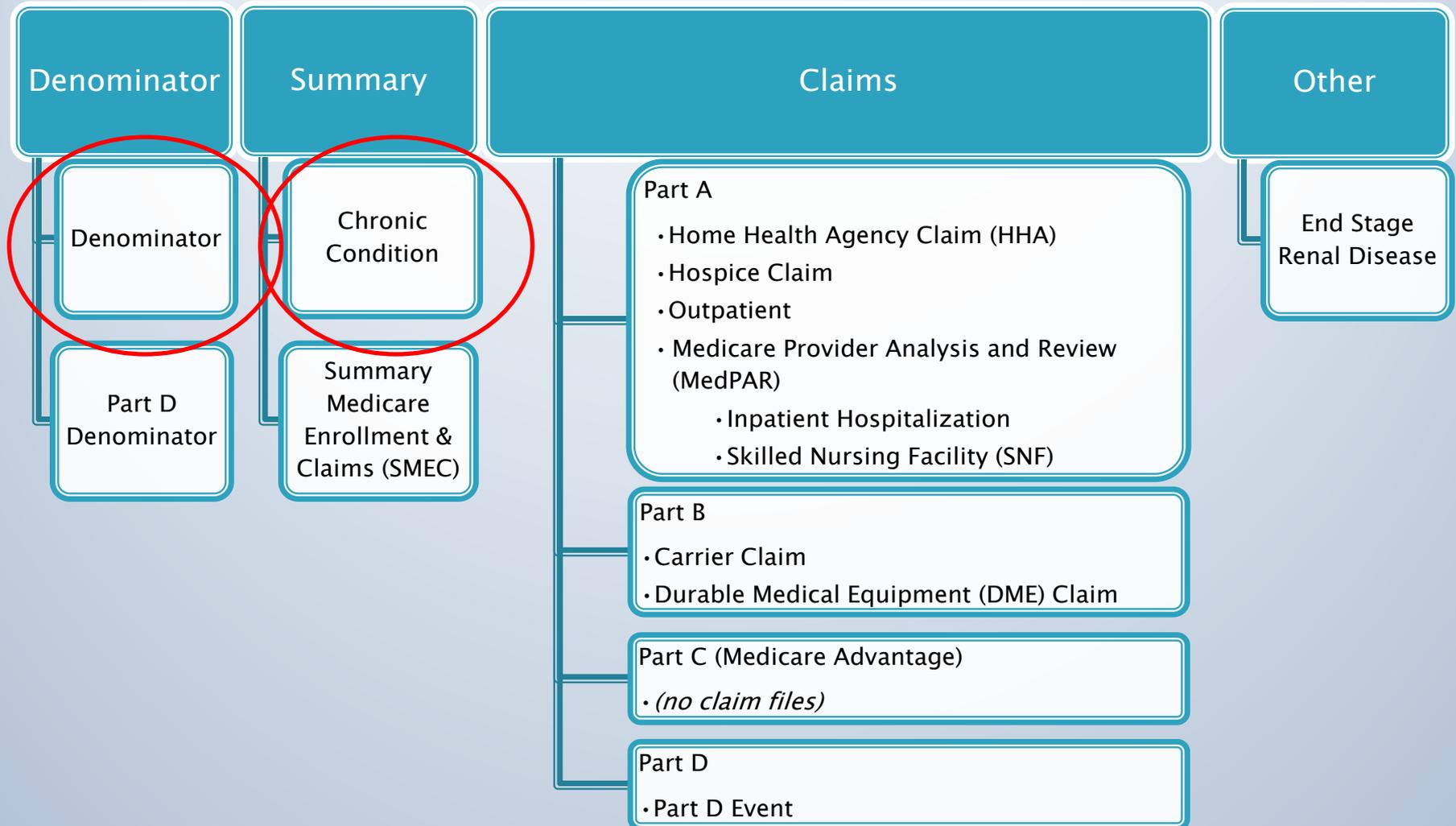 - Linkage ineligibility
  - Reweighting
- Child Survey Participants

Medicare Program Eligibility

- Not everyone is eligible for Medicare
- 10 years Medicare-covered employment and age >=65

> <65 but disability
>
> <65 but End-Stage Renal disease

- Important to remember that program ineligibility is not a linkage ineligibility!

Structure of CMS Medicare Files

Denominator
- Denominator
- Part D Denominator

Summary
- Chronic Condition
- Summary Medicare Enrollment & Claims (SMEC)

Claims

Part A
- Home Health Agency Claim (HHA)
- Hospice Claim
- Outpatient
- Medicare Provider Analysis and Review (MedPAR)
  - Inpatient Hospitalization
  - Skilled Nursing Facility (SNF)

Part B
- Carrier Claim
- Durable Medical Equipment (DME) Claim

Part C (Medicare Advantage)
- *(no claim files)*

Part D
- Part D Event

Other
- End Stage Renal Disease

15

Structure of CMS Medicare Files

Denominator

- Denominator
- Part D Denominator

Summary

- Chronic Condition
- Summary Medicare Enrollment & Claims (SMEC)

Claims

Part A
- Home Health Agency Claim (HHA)
- Hospice Claim
- Outpatient
- Medicare Provider Analysis and Review (MedPAR)
  - Inpatient Hospitalization
  - Skilled Nursing Facility (SNF)

Part B
- Carrier Claim
- Durable Medical Equipment (DME) Claim

Part C (Medicare Advantage)
- *(no claim files)*

Part D
- Part D Event

Other

- End Stage Renal Disease

Linked Data Considerations

- Defining Study Sample
  - Timing of survey data and administrative claims
  - Mortality
  - Program ineligible
- Data Issues
  - *Ever* Variable in Chronic Condition (CC) Summary File
  - Medicare Part C
- Incomplete Linkage
  - Linkage ineligible
  - Reweighting
- Child Survey Participants

Chronic Conditions Summary

- Medicare File, available for 2005-2007
- 21 conditions
- Mid-year, end of year and "ever" flags
- Ever flag includes date first recorded (after 1999)

Chronic Conditions Summary

- Data Considerations
  - Ever flags with date of 1999 may include some people who had claims prior to 1999

Year of First Diabetes Record in CC Summary File 2005

Linked Data Considerations

- Defining Study Sample
  - Timing of survey data and administrative claims
  - Mortality
  - Program ineligible
- Data Issues
  - *Ever* Variable in Chronic Condition (CC) Summary File
  - Medicare Part C
- Incomplete Linkage
  - Linkage ineligible
  - Reweighting
- Child Survey Participants

Linked Data Considerations

- Defining Study Sample
  - Timing of survey data and administrative claims
  - Mortality
  - Program ineligible
- Data Issues
  - *Ever* Variable in Chronic Condition (CC) Summary File
  - Medicare Part C
- Incomplete Linkage
  - Linkage ineligible
  - Reweighting
- Child Survey Participants

Linkage ineligible

- "Linkage ineligible": Survey respondents who are ineligible to be linked.

- "Program ineligible": Respondents who are not part of the administrative program.

Reweighting

- Proc WTADJUST in SUDAAN
- Statistical weights adjusted for linkage ineligibility
- All percentages in this talk use reweighted sample weights
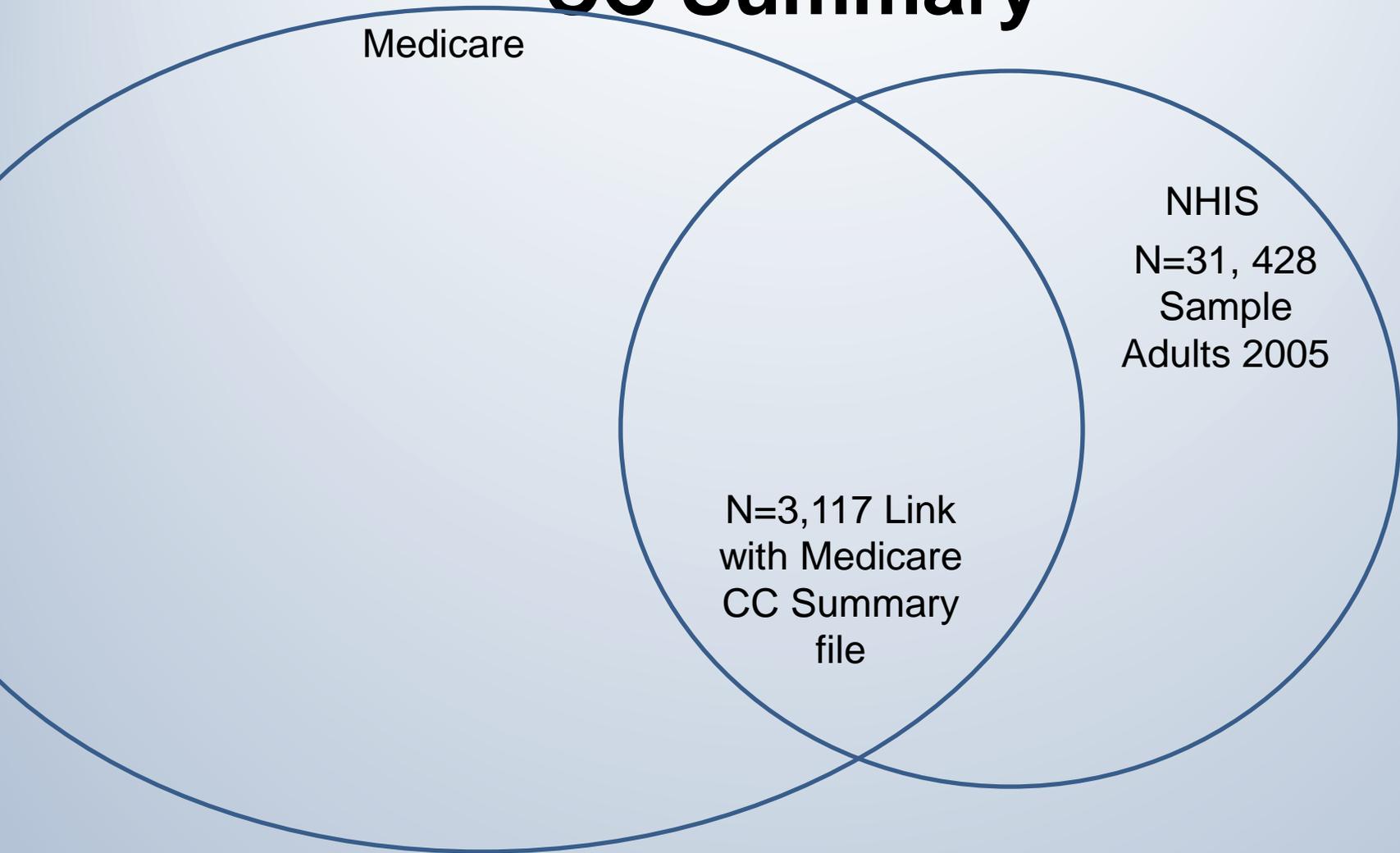
Linked Data Considerations

- **Defining Study Sample**
  - Timing of survey data and administrative claims
  - Mortality
  - Program ineligible
- **Data Issues**
  - *Ever* Variable in Chronic Condition (CC) Summary File
  - Medicare Part C
- **Incomplete Linkage**
  - Linkage ineligible
  - Reweighting
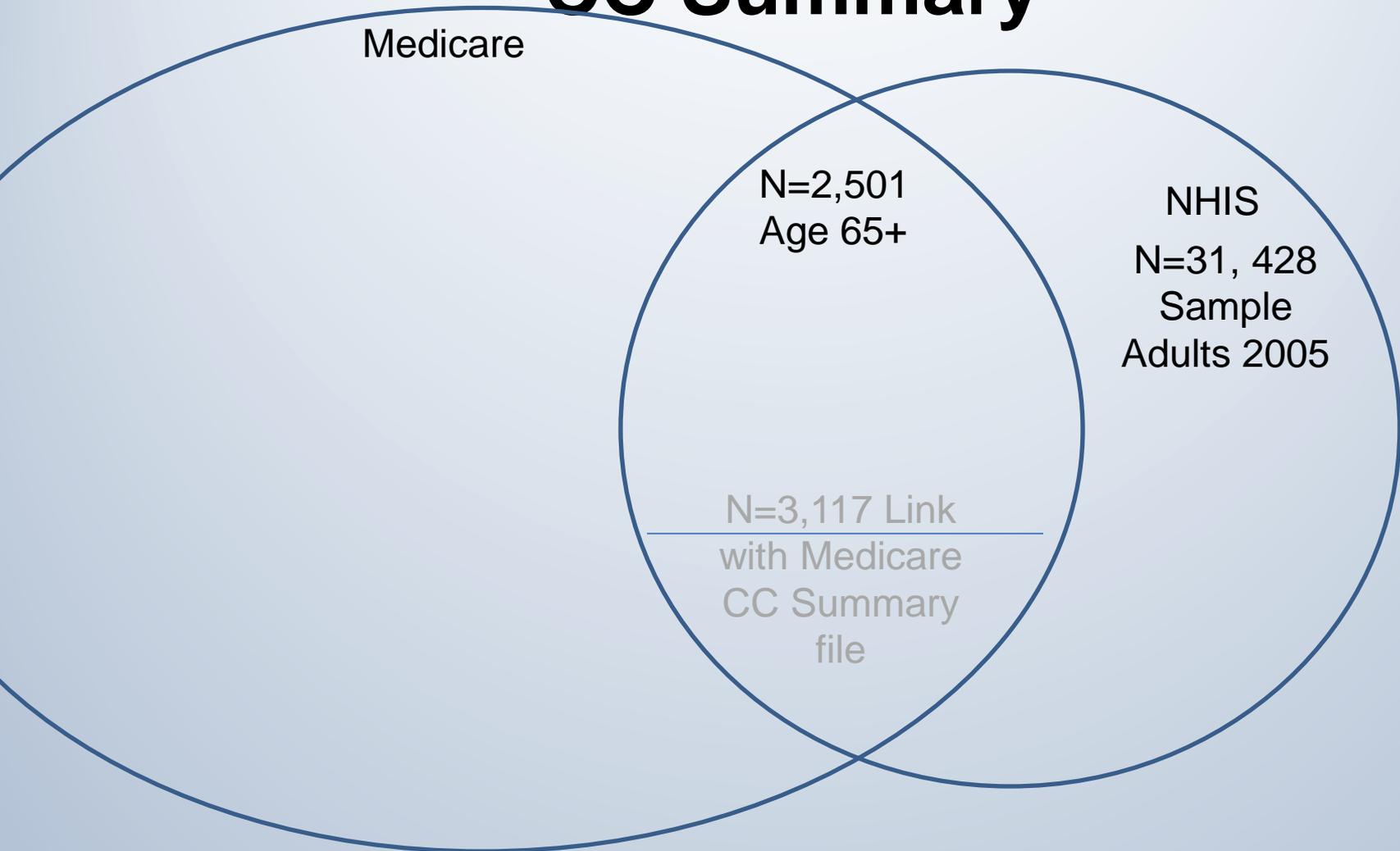- **Child Survey Participants**

Motivating Example

- How does self-report of diabetes in NHIS 2005 compare to the 2005 Chronic Condition Summary File diabetes flag based on Medicare claims?
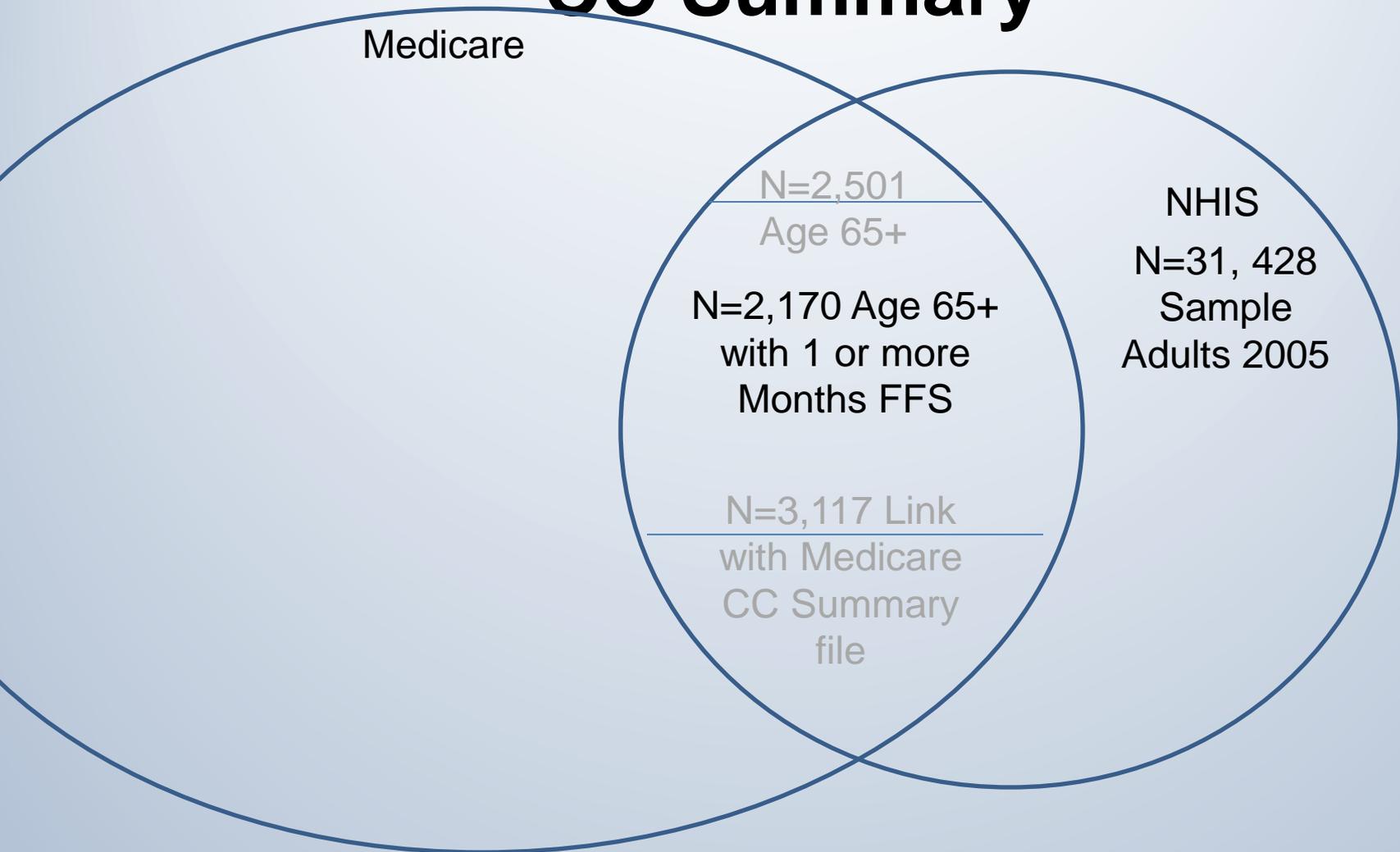
2005 NHIS Sample Adults linked to 2005 CC Summary

Medicare

NHIS

N=31, 428 Sample Adults 2005

N=3,117 Link with Medicare CC Summary file

- However, 22% of the 2005 sample adults who linked with Medicare CC Summary File 2005 were <65

2005 NHIS Sample Adults linked to 2005 CC Summary

Medicare

N=2,501
Age 65+

NHIS

N=31, 428
Sample
Adults 2005

N=3,117 Link
with Medicare
CC Summary
file

- 14.6 % of age 65 or older had no fee for service claims

2005 NHIS Sample Adults linked to 2005 CC Summary

Medicare

N=2,501
Age 65+

N=2,170 Age 65+
with 1 or more
Months FFS

NHIS

N=31, 428
Sample
Adults 2005

N=3,117 Link
with Medicare
CC Summary
file

- NHIS-Medicare linked Sample Adults 2005 age 65 or older at the time of interview, with fee for service claims

NHIS 2005

- Other than during pregnancy, have you EVER been told by a doctor or health professional that you have diabetes or sugar diabetes?
  - Yes: N=420
  - No: N=1691
  - Borderline 2.4%
  - Refused <1%
  - Don't know <1%

Comparison of self-reported diabetes and Medicare claims

Medicare Chronic Condition Summary File	NHIS	
	Self-report Diabetes N	No self-report Diabetes N
Diabetes	390	192
No Diabetes	30	1499

- 2005 NHIS linked to 2005 CC Summary file.
- Adults age 65 and older with one or more months FFS in 2005

Self-reported and Medicare claims for diabetes

	Self-report Diabetes in NHIS N (Weighted %)	No self-report Diabetes in NHIS N (Weighted %)
Diabetes in Medicare	390	192
No Diabetes in Medicare	30	1499

- 2005 NHIS linked to 2005 CC Summary file.
- Adults age 65 and older with one or more months FFS in 2005

Self-reported and Medicare claims for diabetes

	Self-report Diabetes in NHIS N (Weighted %)	No self-report Diabetes in NHIS N (Weighted %)
Diabetes in Medicare	390 *Weighted Row %: 66.1%* *Weighted Column %: 93.0%*	192
No Diabetes in Medicare	30	1499 *Weighted Row %: 98.0%* *Weighted Column %: 87.7%*

- The Medicare CC Summary variable identifies 93.0% of people who self-reported diabetes
  - Column %, similar to sensitivity
- However, if Medicare record is positive, there is only a 66.1% probability a respondent self-reported diabetes
  - Row %, similar to positive predictive value

Self-reported and Medicare claims for diabetes

	Self-report Diabetes in NHIS N (Weighted %)	No self-report Diabetes in NHIS N (Weighted %)
Diabetes in Medicare	390 *Weighted Row %: 66.1%* *Weighted Column %: 93.0%*	192
No Diabetes in Medicare	30	1499 *Weighted Row %: 98.0%* *Weighted Column %: 87.7%*

- Medicare record identified 87.7% of people who self-report no disease
  - Column %, similar to specificity
- If Medicare record is negative, probability a respondent did not self-reported diabetes: Row 98.0%
  - Row %, similar to negative predictive value

Conclusions

- The majority of people who self-reported diabetes in NHIS 2005 had a chronic conditions indicator for diabetes in the Medicare CC Summary file 2005

- If Medicare record is positive, probability a respondent self-reported diabetes 66.5%

- The majority of people who self-reported no diabetes in NHIS 2005 did not have a CC Summary indicator for diabetes in 2005

34

Conclusions

- Further work is needed to:
  - Identify how these factors vary by demographic characteristics
  - Examine potential reasons for greater numbers of Diabetes Ever in the Medicare CC Summary file compared to the self-report
- Important to keep in mind differences between Medicare CC Summary file and self-report
  - Remember Medicare data is collected for billing purposes, not research!

Conclusions

- **Defining Study Sample**
  - Timing of survey data and administrative claims
  - Mortality
  - Program ineligible
- **Data Issues**
  - Ever Variable in CC Summary File
  - Medicare Part C
- **Incomplete Linkage**
  - Linkage ineligible
  - Reweighting
- **Child Survey Participants**

Questions?

http://www.cdc.gov/nchs/data_access/data_linkage_activities.htm
datalinkage@cdc.gov

21 Conditions

- Acute Myocardial Infarction
- Alzheimer's Disease
- Alzheimer's Disease, Related Disorders, or Senile Dementia
- Atrial Fibrillation
- Cataract
- Chronic Kidney Disease
- Chronic Obstructive Pulmonary Disease
- Depression
- Diabetes
- Glaucoma

- Heart Failure
- Hip/Pelvic Fracture
- Ischemic Heart Disease
- Osteoporosis
- Rheumatoid arthritis / Osteoarthritis (RA/OA)
- Stroke / Transient Ischemic Attack
- Breast Cancer
- Colorectal Cancer
- Prostate Cancer
- Lung Cancer
- Endometrial Cancer

- Linkage Rates

Table 2. Sample Size and Unweighted Linkage Rates for Linkages of NCHS Surveys Containing only 1999-2007 Medicare Enrollment and Claims Data by Survey and Age Group[1]
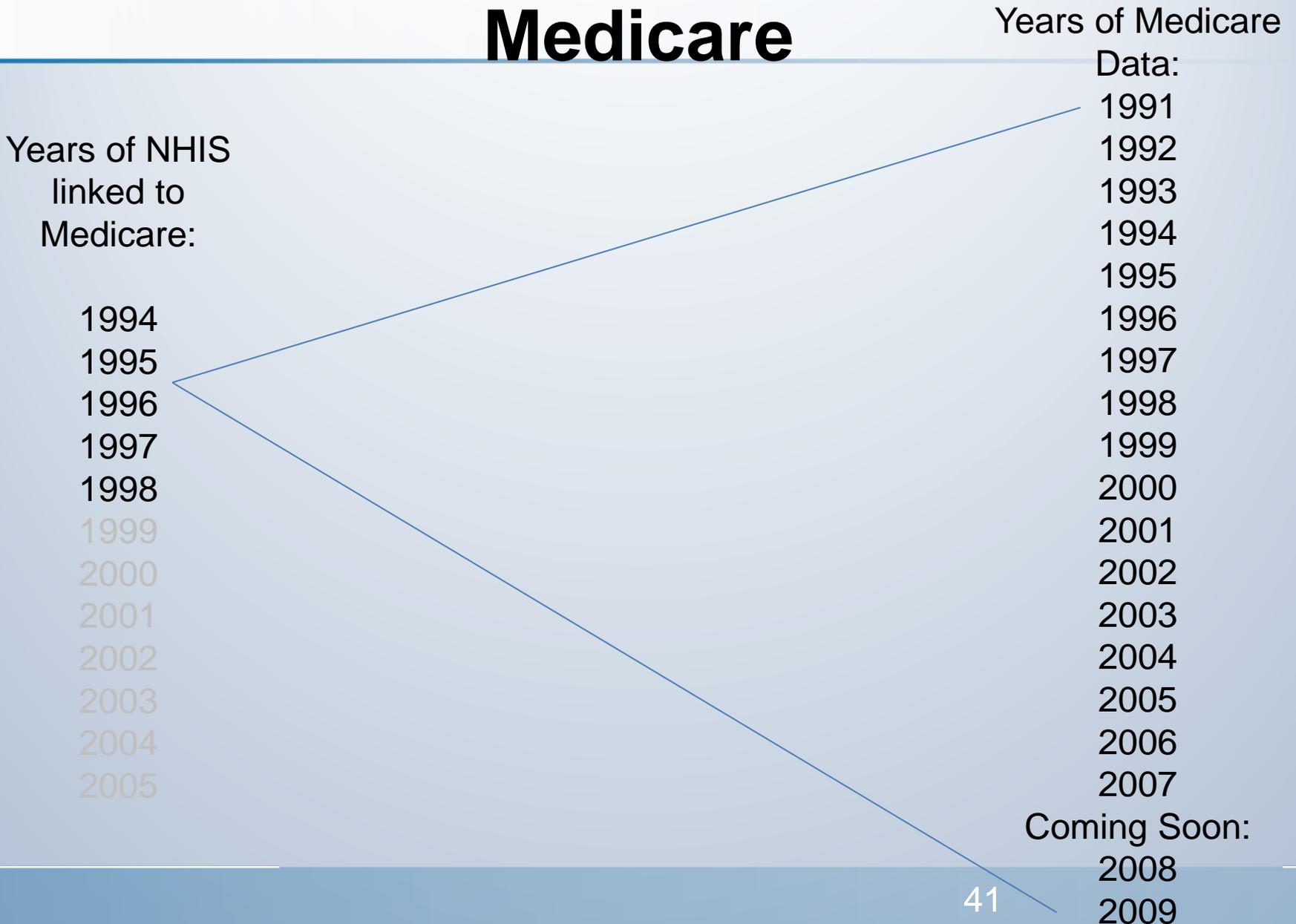
	Total Person Sample (Column 1)	Sample Eligible for the 1999-2007 Medicare Linkage[2] (Column 2)	Number of Respondents with Information on any Medicare Denominator File[3] (Column 3)	Linkage Rate for Total Sample (Column 3/Column 1)	Linkage Rate for Eligible Sample (Column 3/Column 2)
NHIS 2005	98,649	44,835	6,926	7.0%	15.4%
<65	85,656	39,075	1,282	1.5%	3.3%
>=65	12,993	5,760	5,644	43.4%	98.0%

- http://www.cdc.gov/nchs/data/datalinkage/ cms_medicare_methods_report_final.pdf

Diabetes Algorithm

Algorithms	Reference Time Period (# of years)	Valid ICD-9/CPT4/HCPCS Codes[1]	Number/Type of Claims to Qualify[2,3]	Exclusions
Diabetes	2 years	DX 249.00, 249.01, 249.10, 249.11, 249.20, 249.21, 249.30, 249.31, 249.40, 249.41, 249.50, 249.51, 249.60, 249.61, 249.70, 249.71, 249.80, 249.81, 249.90, 249.91, 250.00, 250.01, 250.02, 250.03, 250.10, 250.11, 250.12, 250.13, 250.20, 250.21, 250.22, 250.23, 250.30, 250.31, 250.32, 250.33, 250.40, 250.41, 250.42, 250.43, 250.50, 250.51, 250.52, 250.53, 250.60, 250.61, 250.62, 250.63, 250.70, 250.71, 250.72, 250.73, 250.80, 250.81, 250.82, 250.83, 250.90, 250.91, 250.92, 250.93, 357.2, 362.01, 362.02, 366.41 (any DX on the claim)	At least 1 inpatient, SNF or HHA claim or 2 HOP or Carrier claims with DX codes during the 2-yr period	

Extra info on early years of NHIS-Medicare

Years of NHIS linked to Medicare:

1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005

Years of Medicare Data:

1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
Coming Soon:
2008
2009

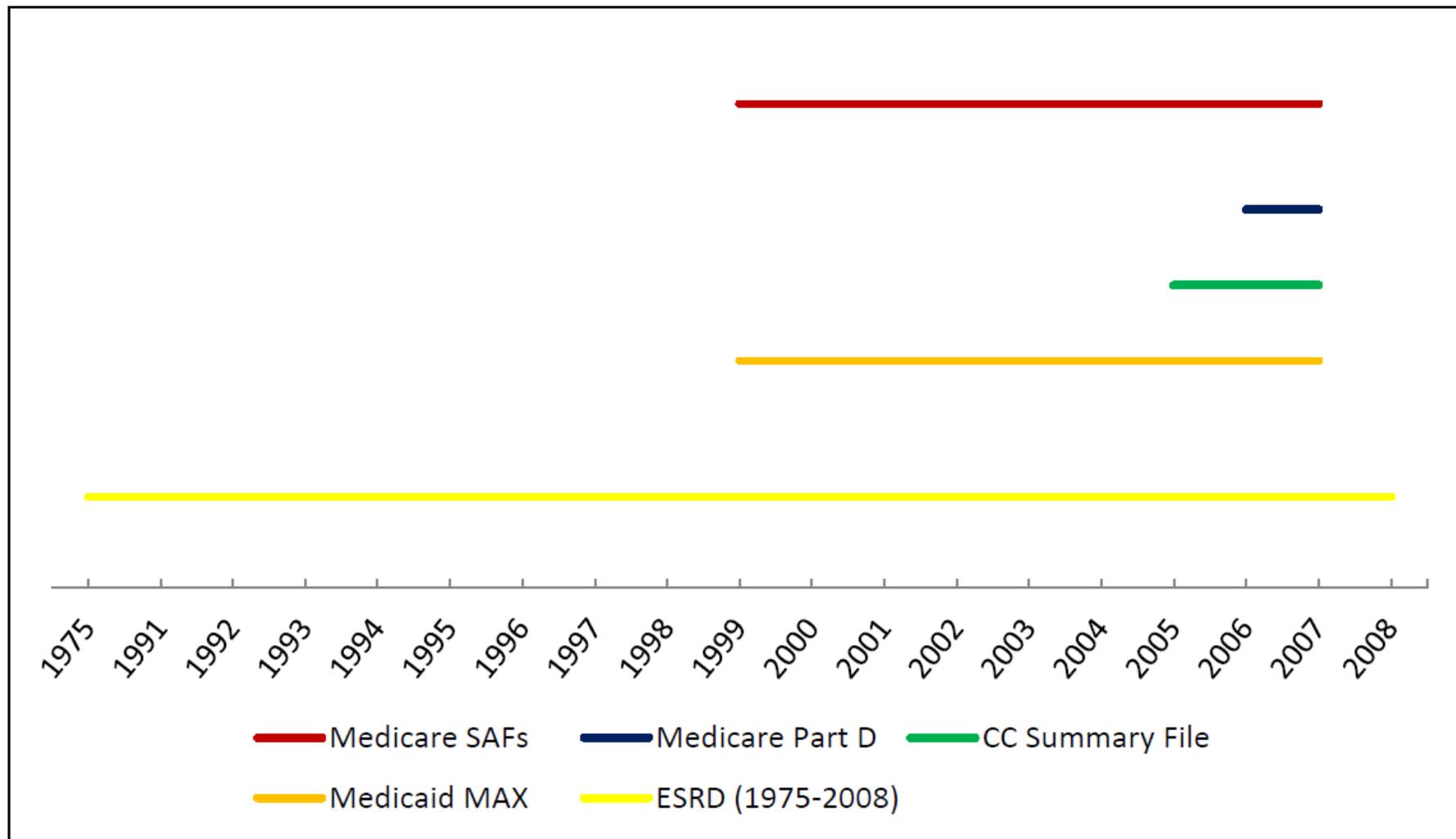Surveys: 1999-2005 NHIS, 1999-2004 NHANES, 2004 NNHS

Table 2. Unweighted percent of NCHS survey participants enrolled in managed care for surveys containing only 1999-2007 Medicare enrollment and claims data by survey and age at interview

	1999	2000	2001	2002	2003	2004	2005	2006	2007
	%								
NHIS 1999	19.8	19.4	17.3	14.9	13.9	14.0	15.1	18.9	21.6
<65	13.1	13.4	11.9	10.8	10.6	11.0	12.7	17.1	20.3
>=65	21.2	21.1	19.3	16.8	15.8	16.1	17.2	20.7	23.2
NHIS 2000	20.6	20.2	17.7	15.7	14.6	15.2	16.2	20.0	23.0
<65	12.6	12.4	11.4	10.1	9.8	11.3	12.0	17.2	20.7
>=65	21.9	21.8	19.5	17.7	16.9	17.5	19.1	22.4	25.3
NHIS 2001	19.5	19.0	16.8	14.5	13.7	13.6	14.8	18.6	21.5
<65	10.4	10.1	9.7	9.0	9.1	9.8	11.4	16.1	19.3
>=65	21.1	20.6	18.4	16.2	15.6	15.6	17.0	20.5	23.4
NHIS 2002	20.7	20.1	17.9	15.6	14.4	14.1	15.3	19.4	21.6
<65	10.9	10.6	9.1	7.5	7.1	8.1	9.4	15.3	17.9
>=65	22.3	21.6	19.5	17.4	16.7	16.5	18.3	21.9	24.5
NHIS 2003	20.1	19.2	17.0	14.8	13.6	13.9	15.3	19.0	21.0
<65	10.1	9.3	8.8	7.4	6.9	7.9	9.8	14.9	17.5
>=65	21.8	20.9	18.4	16.2	15.2	15.8	17.5	21.1	23.2
NHIS 2004	19.8	18.9	17.2	14.7	14.0	14.0	15.4	19.5	22.8
<65	9.9	8.6	7.9	6.8	6.8	6.8	8.8	13.8	18.6
>=65	21.2	20.4	18.7	16.0	15.3	15.6	17.4	21.8	25.0
NHIS 2005	18.8	18.4	16.3	14.1	13.2	13.2	14.8	18.6	21.7
<65	8.6	9.1	7.4	5.6	5.3	5.7	8.5	13.5	17.1
>=65	20.4	19.9	17.9	15.6	14.8	14.8	16.5	20.4	23.8

Exclusion of claims paid by a source other than Medicare (e.g., Medicare Part C plans)

CMS generally does not receive claims data for Medicare beneficiaries who enroll in Medicare Part C plans (including private fee-for-service plans paid on a capitation basis). Please note that exceptions to this do exist. For example, all Hospice claims are processed as Medicare claims regardless of whether the beneficiary is in a Fee for Service (FFS) or a Medicare Part C plan. During the time covered by the linked database, enrollment in Medicare Part C plans increased from approximately 6% of beneficiaries in 1991 to 20% in 2007.

In general, studies based on analysis of claims data should exclude Medicare Part C enrollees from their beneficiary samples. For health outcome or epidemiologic studies (as opposed to utilization or cost studies) an alternative approach for dealing with Medicare Part C enrollees is to include them for the time period prior to entering a Medicare Part C plan and then censor them at the time they enter a Medicare Part C plan.

As noted above the Medicare Denominator file can be used to identify Medicare Part C enrollees. A summary of the percent of NCHS survey respondents who were enrolled in a Medicare Part C plan by year and survey can be found at http://www.cdc.gov/NCHS/data/datalinkage/nchs-cms_medicare_linked_data_managed_care_enrollment_tables.pdf.

Researchers should consider the percent of respondents enrolled in a Medicare Part C program when determining the feasibility and sample sizes of their proposed research projects.

The following documents and citations provide detailed information about Medicare Part C enrollees and the Medicare Utilization Files and how to address them in analyses:

http://www.resdac.umn.edu/Tools/TBs/TN-009_MedicareManagedCareEnrolleesandUtilFiles_508.pdf

Virnig BA et al. Survival analysis using Medicare data: example and methods. *Health Services Research* 2000 Dec; 35(5 Pt 3):86-101.

Medicare Part C plans are also referred to as Medicare Advantage (MA) and include Health Maintenance Organizations (HMOs), Preferred Page **3** of **16**

Provider Organizations (PPOs), Private Fee-for-Service (PFFS) Plans, Special Needs Plans, and Medicare Medical Savings Account Plans.

Extra information on Borderline Diabetes

	Self-report Diabetes in NHIS	No self-report Diabetes in NHIS	Borderline Diabetes in NHIS
Diabetes in Medicare	343	83	31
No Diabetes in Medicare	110	1779	27