ResDAC
RESEARCH DATA ASSISTANCE CENTER

Sources and Use of Medicare Enrollment Information

THE MASTER BENEFICIARY SUMMARY FILE

Beth Virnig, Ph.D.
Associate Dean for Research and Professor
University of Minnesota

In the beginning…

- There was a single denominator file, it was created for researchers, it was thin (80 columns) and we suggested that everyone get a copy of it for every study

Now…

- There are multiple different ‘denominator files’
  - The Beneficiary Summary file (includes Part D Denominator)
  - The CMS Denominator
  - The Part D Denominator
  - The PedSF and SumDenom (for SEER/Medicare)
- And we have every reason to expect that denominators will continue to evolve (and perhaps multiply)
- Because now we have the Master Beneficiary Summary File with 4 segments

Master Beneficiary Summary File

SEGMENTS

- There are 4 segments to the new Master Beneficiary Summary File
  - Beneficiary Summary File (A/B/C/D)
  - Chronic Conditions
  - Cost & Utilization
  - NDI Death Information (includes ICD-10 Cause of Death)

- This presentation will be discussing the Beneficiary Summary File segment

So, thinking about denominators conceptually offers advantages

- The reasons for a ‘denominator’ do not change
- The specific details are well documented and explained
- Conceptual understanding is essential to proper study design and effective use of the data

Contents of Denominators:

- Variables used for patient identification
- Variables used for demographic information
- Variables used to track eligibility for receiving particular services under Medicare

Denominators

- Recommended source of demographic variables for Medicare analysis:
  - Date of birth/Age
  - Date of death
  - Sex
  - Race

Who is included in the Denominator file or the Beneficiary Summary file?

- Annual file containing all beneficiaries *enrolled* for even one day in the CY
  - The file isn't limited to users unless you do so by selection
- Eligibility is determined by SSA & RRB based on information from SSA & RRB
- All benefit groups - Unless you specify otherwise
- No specific indicator for 'new beneficiaries'

So, why are all these files challenging to use?

- DATE STAMPING
  - Date stamping is the idea that these identifying/classifying variables that can or do change over time are still only represented once in the file.
  - Understanding the rules about which value is contained is essential for interpreting the information
  - The most common options for date stamping are:
    - » The first value
    - » The last value
    - » The value that was noted when the file was created

Sources of Denominator Data

- CMS
- Social Security Administration (SSA)
- Railroad Board (RRB)
- States
- *Claims*
- *Managed Care Organizations*

Underlying all Denominator Files is:

- CMS Enrollment Database (EDB)
  - CMS takes the data from all these sources and stores them in their own database called the Enrollment Database (EDB)
- The EDB contains eligibility and enrollment information for every beneficiary ever entitled to Medicare
- Once a year, data are extracted from the EDB to create the CMS Denominator File
- CCW Beneficiary Summary File is updated for 1 full year

HIC—Medicare's Unique Identifier

- 11 digit identifier
  - 9 digit CAN (claim account number–usually SSN under which benefits are claimed)
  - 2 digit BIC–beneficiary identification code allows for beneficiaries sharing same SSN (or RRB ID) to be distinguished

The IDs used by CMS for Medicare Users have not changed…

- The IDs researchers receive have changed dramatically.
- The Actual IDs that beneficiaries use is called the HIC (Health Insurance Claim Number)
- Research files now contain the BeneID, which is not the actual Beneficiary ID but is a unique, study-specific ID (more on that in a minute)
- However, the underlying properties of the HICs deserve some attention…

The HIC is based on the SSN.

- The SSN is not a totally random number
- First 3 digits–state in which the SSN was assigned or state of residence at the time the SSN was obtained
- Next 2 digits–group number–sequencing number used by SSA
- Last 4 digits randomly assigned
  - This property is used for efficient sampling of Medicare beneficiaries
  - A systematic sample of a random number is a random sample

BIC

- Assigned by Social Security Administration to explain the reason for claiming benefits under a particular work history (i.e., SSN).
- No two people claiming benefits under the same SSN can have the same BIC
- The SSA has over 60 categories of BICs that reflect both justification for benefits and level of benefit.

Facts about HICs, BICs, SSNs and BeneIDs

- The HIC is unique. No two people ever share the same HIC – either current or historical
- Multiple persons can claim Medicare benefits under the same SSN (work history). The addition of the BIC results in a unique identifier for each person

Facts (continued)

- Even though people can share an SSN, most people now have their own SSNs. HICs are assigned based upon the SSN that is used to claim benefits. A person may have their own SSN but claim benefits under their spouse's work history. SSN benefits are generally assigned to maximize retirement payment.

Facts (continued)

- HICs can change. While HICs are generally stable, people do change HICs on occasion. This is typically the result of the decision to claim benefits under a different work history, and is often tied to SSA payments rather than Medicare benefits. For all research files, CMS automatically links beneficiaries over time even if they change their HICs.

Facts (continued)

- The BeneID is uniquely assigned for each study. Different studies will have records with the same BeneIDs. These are not the same person, they are the same study-specific ID. They cannot be used to combine your data with your colleague's.

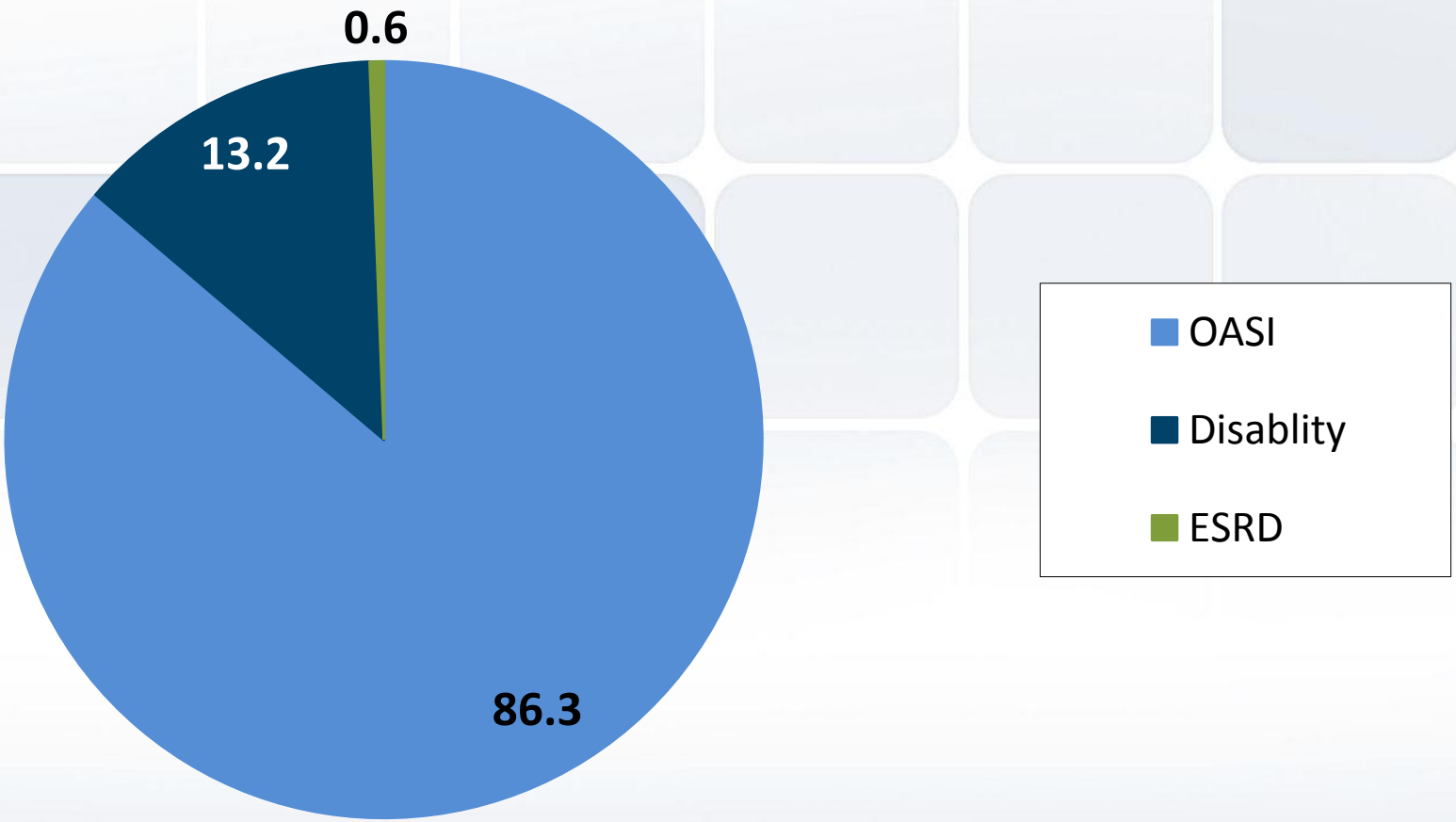
The world is better

- A cross-walk between the BeneID and the HIC will be available for people who need the ability to link back to some data that contained actual HICs such as for longitudinal studies or linking with an outside data source.
- Consider this change to be major improvement. Because there is NO analytic value in the HIC, a random number is equally valid and significantly reduces the risk associated accidental security breaches.

Residency

- State, county and ZIP code of residence are the mailing address for official correspondence
- Some persons have their mail sent to another person (e.g., son, daughter, guardian)
- Analyses comparing state of treatment with state of residency generally show high concordance
- Residency is:
  - based on the information available when the record is finalized for the Denominator file (so it may reflect changes that happen after the end of the CY).
  - Beneficiary Summary file residence reflects information as of 12/31/XX (CY of file)

Medicare Beneficiaries



Entitlement

- Original entitlement
  - old age
  - disability
  - ESRD
  - disability+ ESRD
- Current entitlement
  - see above categories

Medicare Status Code

- Medicare Status Code (MSC) combines current entitlement and ESRD
  - 10 aged w/out ESRD
  - 11 aged w/ ESRD
  - 20 Disabled w/out ESRD
  - 21 Disabled w/ ESRD
  - 31 ESRD only

MSC is important because the beneficiaries in each of the 3 programs are not the same

	Elderly	Disabled	ESRD
% male	41.6%	55.6%	54.5%
Annual mortality	6.1%	2.6%	8.1%
Mean age	74.6 years	49 years	46 years
Top DRG for inpatient care	Heart failure	Psychoses	Vascular procedures (e.g., for dialysis)

Age and Date of Birth

- Age is calculated differently for the Denominator and Beneficiary Summary file.
  - In the Denominator AGE is the YOUNGEST the person will be.
    - » People turning 65 will be listed as 64 in the file
  - In the Beneficiary Summary File, AGE is the OLDEST they COULD be (age at the end of the calendar year—regardless of whether they survived).
- Date of birth is the actual beneficiary DOB.

Really, really old people

- There are persistent concerns that some deaths are missed by the Medicare program (or SSA). The frequency of ‘really, really old people’, that is people over age 90, 100 or 120, is greater in Medicare than the census.

Really, really old people: Medicare vs. Census (2006 data)

	Medicare	Census
90-94	1,252,640	1,196,000
95-99	314,880	369,000
100+	177,620	68,000
100-119	146,100	n/a
120-129	26,340	n/a
130+	5180	n/a

Really, really old people (continued)

- There is no standard way to remove these people or even consistent practices regarding removing such people.
- Options:
  - anyone over 100 (or 90) who has NO health care use in a year be deleted.
  - Anyone over 90 who has no Part B coverage be deleted
  - Anyone older than the oldest person in the US be deleted
- This is still a really small number of people relative to the total Medicare population (.33% are 100 or over)

Sex

- Sex is coded 1=male 2=female
- There are no missing values for this field
- Persons with missing information have it filled according to the rule: if age is less than 65 and sex missing then sex=male if age is greater than or equal to 65 and sex is missing then sex=female

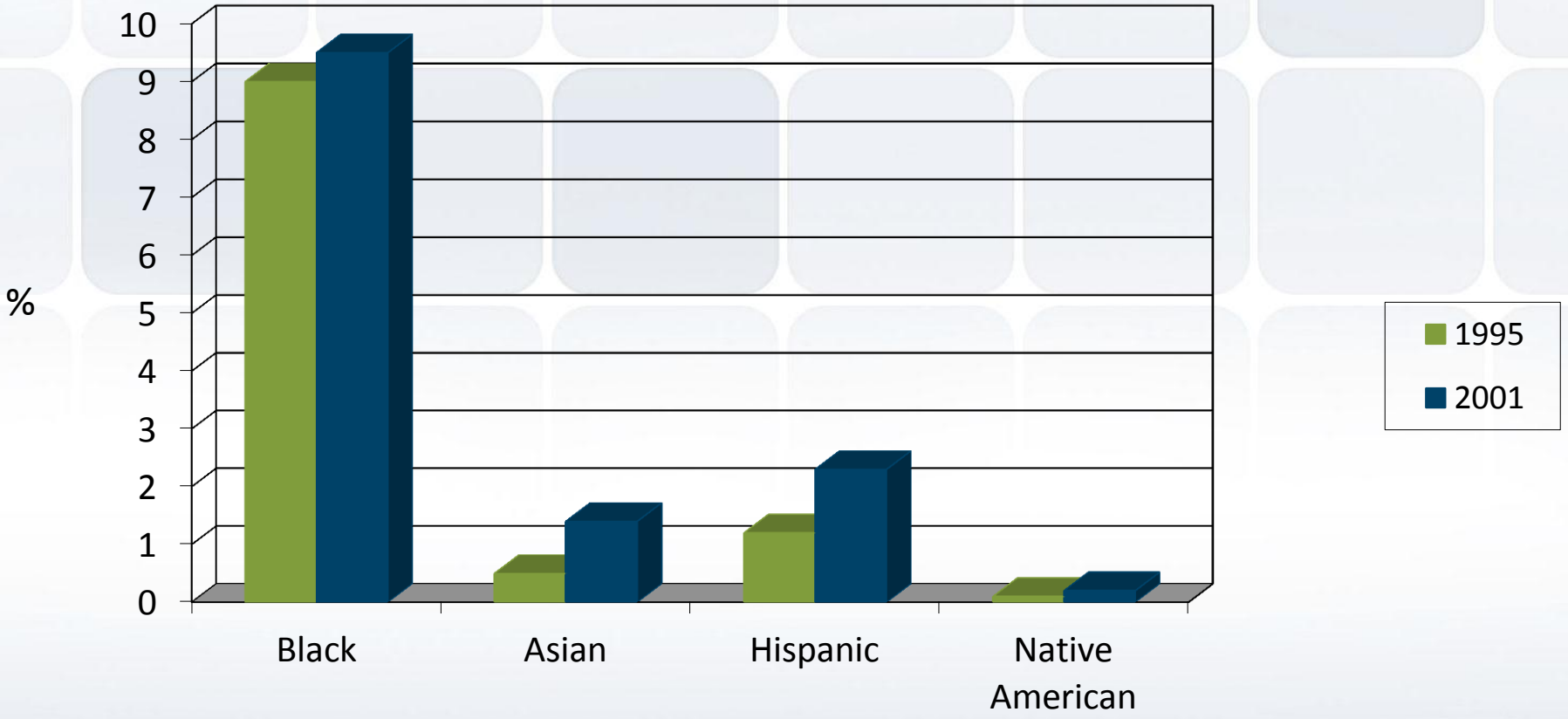
Signs of potentially mistaken sex assignment

	% male	% female
Prostate cancer	100	0
Ovarian or cervical cancer	0.02	99.98
Breast reconstruction surgery	0	100

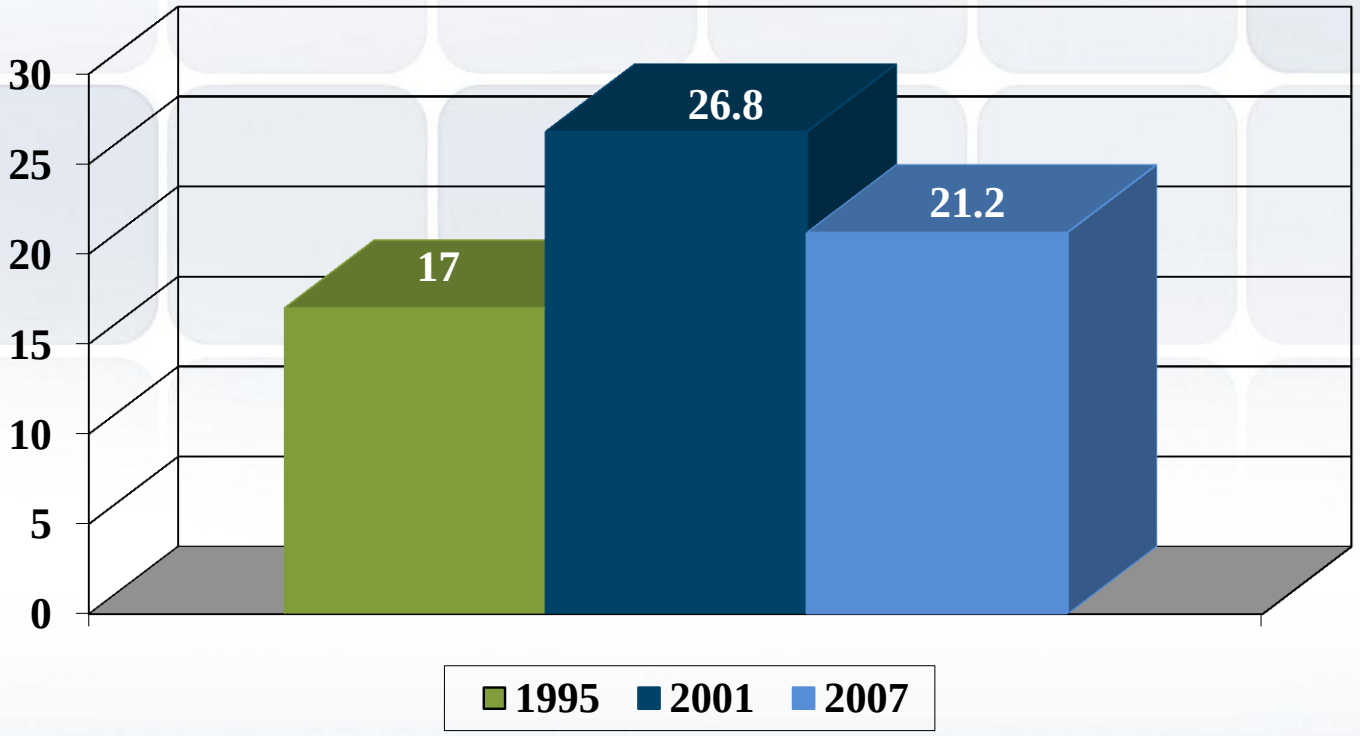
Race

- Previously coded as:
  - white, black, other, unknown
- Effective 1994, race codes were expanded to:
  - white, black, Asian, Hispanic, Native American, other, unknown
- Efforts to update racial classification of beneficiaries are ongoing
- The Hispanic race/ethnicity code has an estimated sensitivity of about 35%
- There is no widely accepted grouping of racial codes

Race codes reflect greater racial diversity

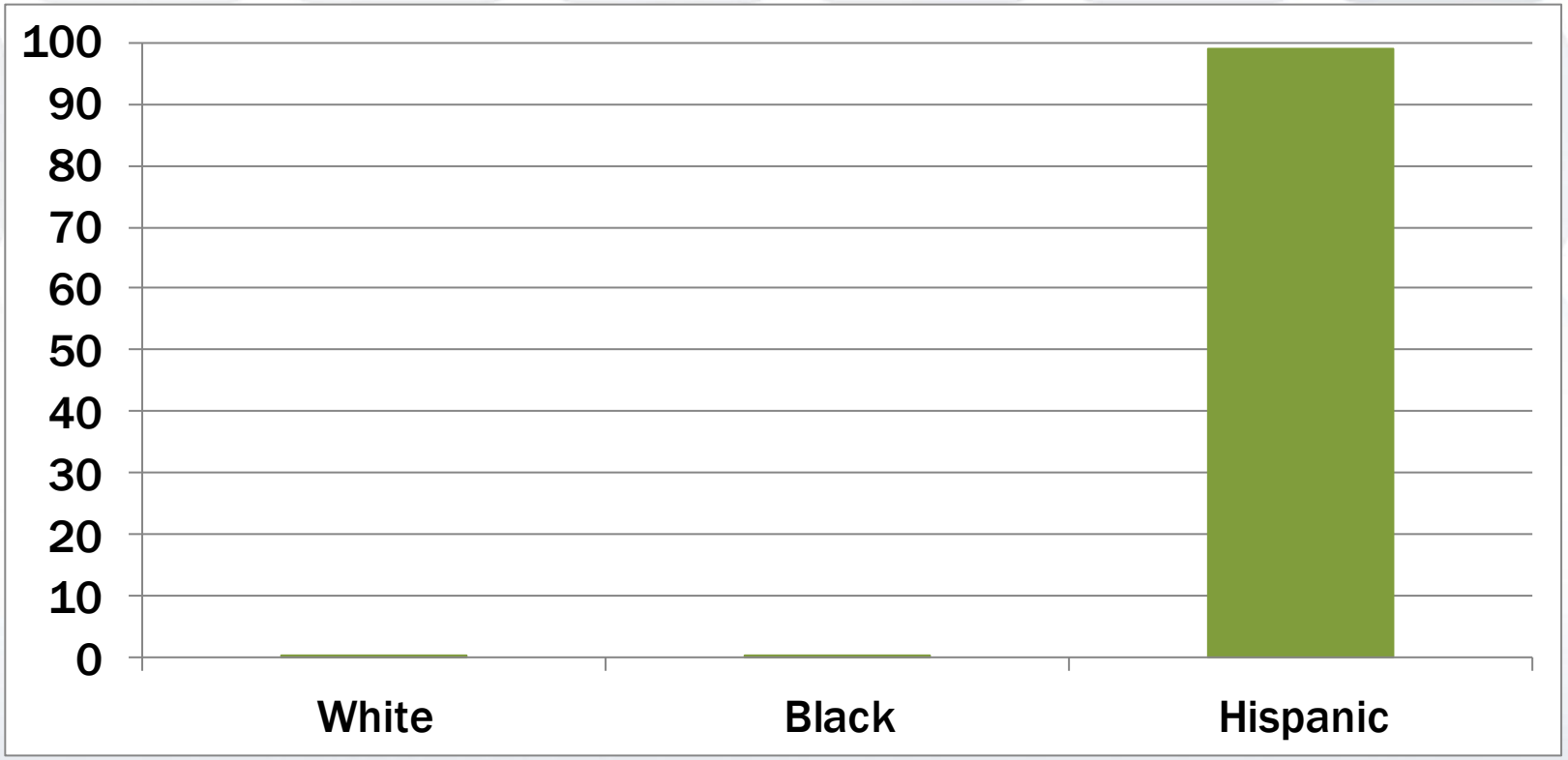


But there is still work to be done:
% Hispanic Race for Residents of Puerto Rico

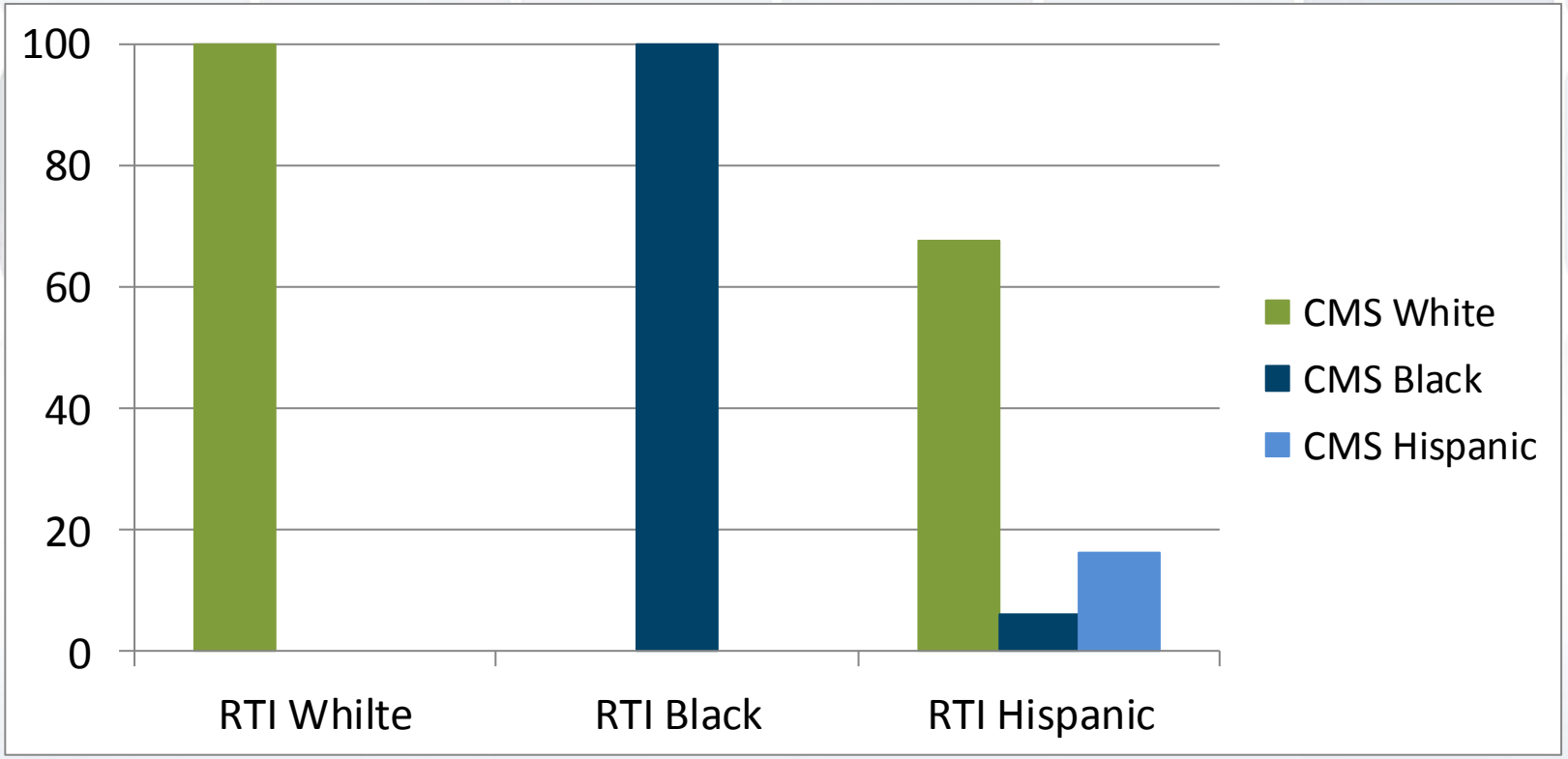


But, the information may not be 'wrong' – beneficiaries are BOTH white and Hispanic or black and Hispanic…

The New RTI Race variable may help a bit because it takes surname information into account… (Puerto Rico again)



So, how do the two variables line up? (again, Puerto Rico)



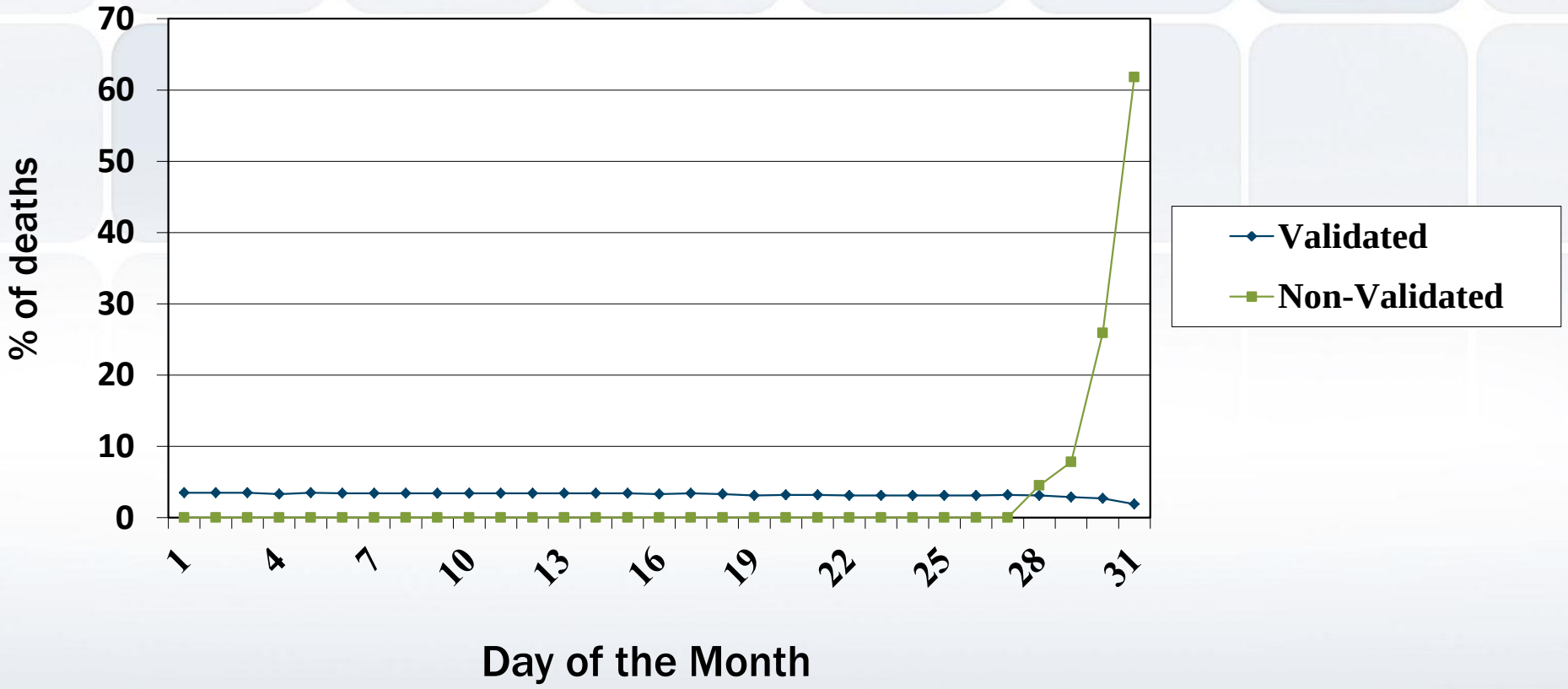
Mortality

- Two fields–date of death and death DATE validation field
- Death dates are missing if the beneficiary is alive and non-missing if they are deceased
- 100% of deaths are validated
- 96% of death DATES validated
- Validated death dates are noted with ‘V’

Date of Death Information

- Social Security Administration
  - Primary source of date of death
- Claims are used to identify beneficiaries who might have died.
  - Gathered from hospitalization claims indicating the patient died in the hospital
  - No beneficiary is determined to be dead without a confirmation process, but report of an in-hospital death will trigger this process

Non-validated death dates are assigned a death date at the end of the month



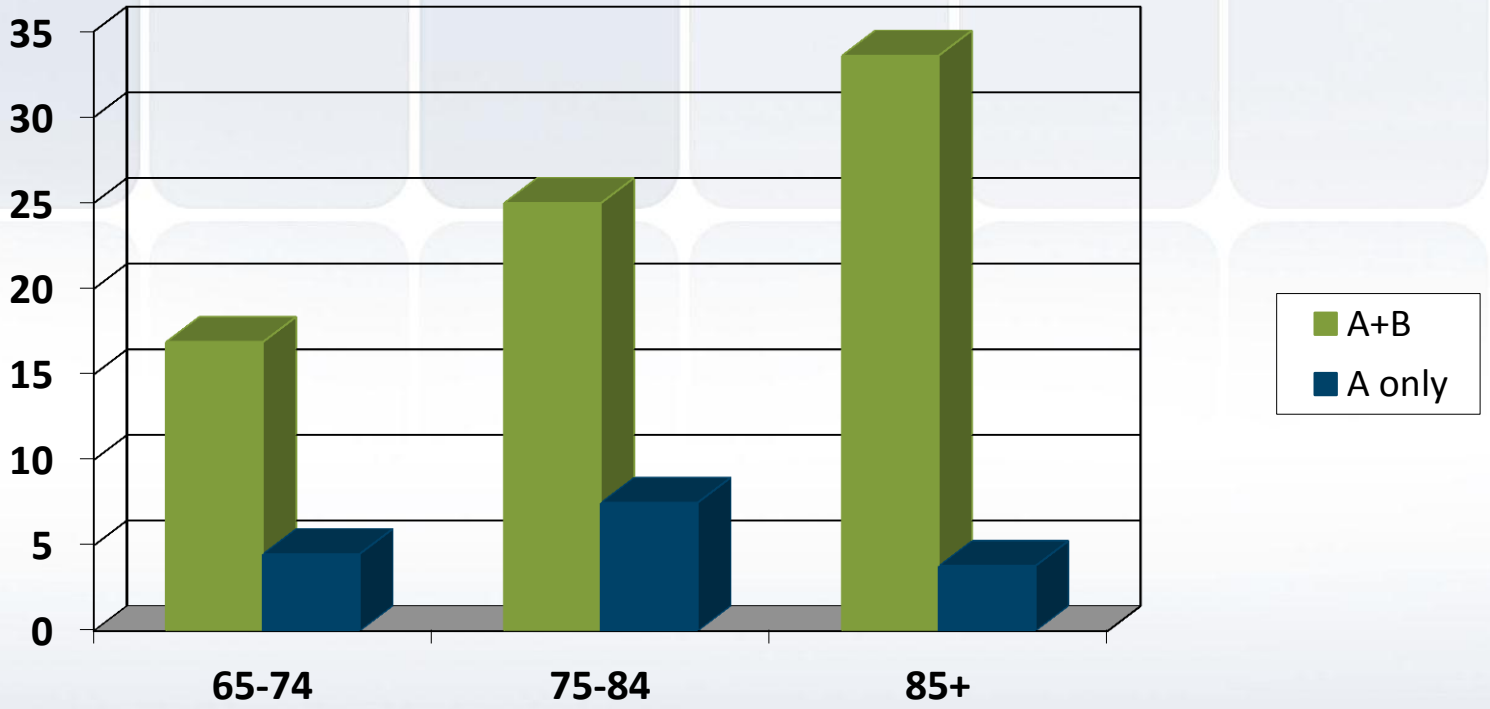
Survival Time and Non-validated Death Dates

- Including non-validated death dates as actual death dates will over-estimate survival times for those individuals

Benefits

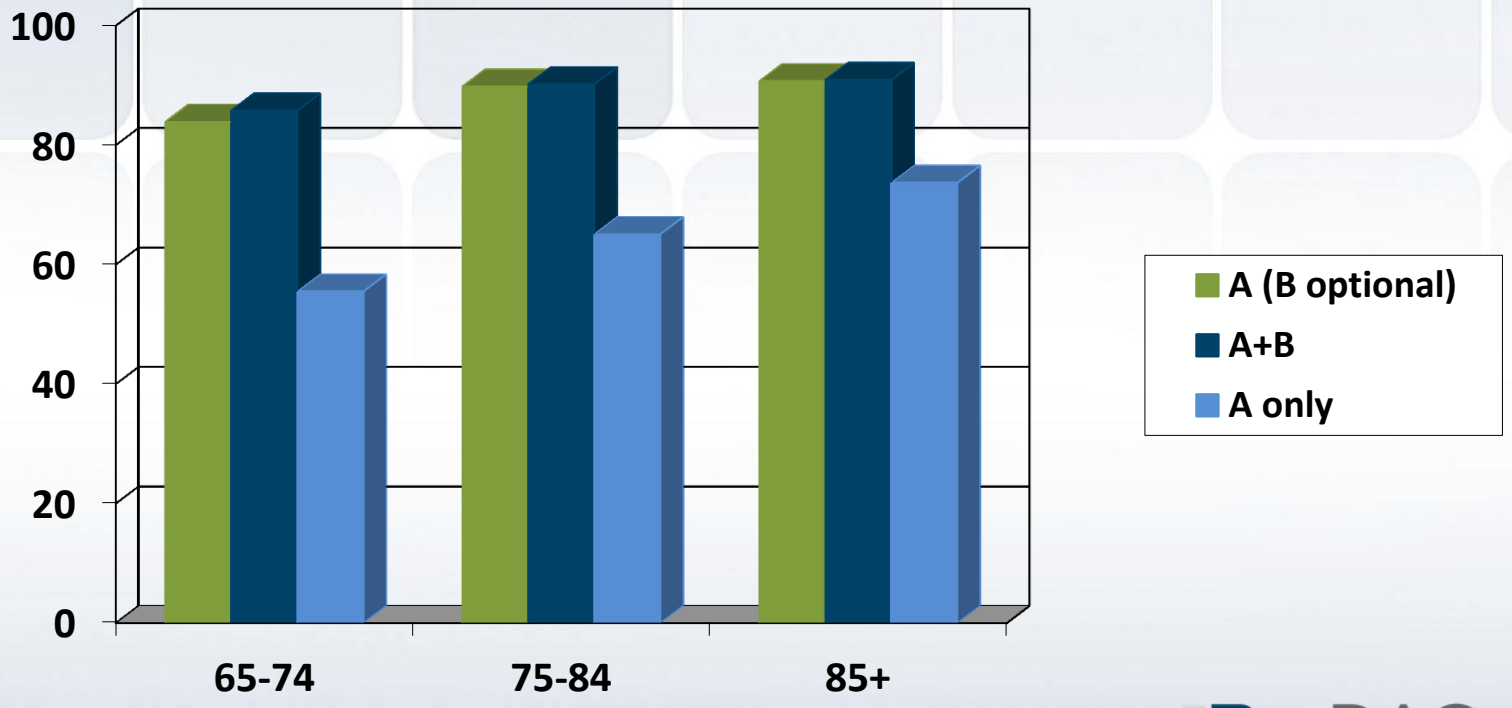
- Part A, Part B
- 94% both Part A and Part B
- 6% have more months Part A services than Part B.
- Beneficiaries are not required to have Part B benefits.
- They can waive Part B benefits and restart without penalty at a later date if they have other health insurance that provides the same coverage.

Examining hospitalization rates by coverage level (per 100 enrollees) supports the conclusion that persons with A-only coverage probably have incomplete claims, even for Part A services



This pattern is seen with patients 'known' to have treatment

(Hospitalization rates for 1992 incident colo-rectal cancer cases treated with surgery—SEER/Medicare)



Medicaid Paying Medicare Premiums

- All states exercise the option of paying Medicare premiums for at least some people
- This can take 3 forms:
  - State pays premiums only (4.6%)
  - State pays premiums and cost sharing (45%)
  - State provides full Medicaid benefits (50.4%)
- This is noted in the Denominator and Beneficiary Summary files as ‘state buy in’

State Buy-In Information in the Denominator File

- Summary count of total months state-buy-in (A or B or both).
- Monthly indicators specifying whether State buy-in covered Part A, Part B or both A and B benefits

Source of State Buy-in Data

- States: When a beneficiary's Part A and/or Part B premiums are paid by the state, the state informs CMS and CMS then bills the state instead of the beneficiary for the Part B premiums.

State Buy-in: What does it tell us?

- The ‘state buy-in’ indicator tells whether a particular beneficiary is covered by one of the three programs (but not which program)
- While it CAN be assumed that persons with state buy in have resources < 2 times the SSI threshold, it CANNOT be assumed that persons without state buy-in have incomes > 2 times the SSI threshold (the indicator is not a clean proxy for income)

Monthly Enrollment Status: Source

- Medicare benefits are determined on a monthly basis
- CMS gathers and maintains information related to each beneficiary's enrollment status, including:
  - Part A enrollment
  - Part B enrollment
  - Disenrollment

Monthly Indicators

- For each month, entitlement/buy-in indicator that summarizes Part A and Part B benefits and state buy-in.
  - Not entitled (0)
  - Part A only (1)
  - Part B only (2)
  - Part A and Part B (3)
  - Part A, State buy-in (A)
  - Part B, State buy-in (B)
  - Parts A and B, State buy-in (C)

Monthly Indicators

- **Examples of actual cases:**
- CCCCCCCCCCCC (12 months, A&B SBI)
- 333333333333 (12 months A&B)
- 333333333333 (12 months A&B)
- 111111333333 (5 mon. A, then 7 mon A&B)
- 111111111111 (12 months A)
- 333300000000 (4 mon A&B,8 mon not elig)
- 000000000033 (10 mon not elig,2 mon A&B)
- 333333330000 (8 mon A&B, 4 mon not elig)

Beginning in 2006, in the Beneficiary Summary file, we get additional information for each month…

- 01: QMB only
- 02: QMB and Medicaid coverage including Rx
- 03: SLMB only
- 04: SLMB and Medicaid coverage including Rx
- 05: QDWI
- 06: Qualifying individuals
- 08: Other Dual eligible (Non-QMB, SLMB, QWID, or QI)with Medicaid coverage including Rx
- 09: Other Dual eligible but without Medicaid coverage

Source of Managed Care Enrollment Data

- Managed Care Organizations (MCOs) Transmit enrollment and disenrollment data as well as enrollment corrections to CMS
- The accuracy of these data is essential to ensure that MCOs are paid a monthly premium from CMS and to make sure that claims that are inadvertently submitted to CMS are rejected.

Managed Care Information for Research

- Monthly HMO indicators and a summary count of Months HMO coverage.
- Summary count does not distinguish across HMO types
- Neither summary count nor monthly indicators include information about the specific plan or switching between plans

Managed Care Enrollment

- Other than for demonstration projects, Managed Care Enrollees are required to have both Part A and Part B benefits
- 99.9 % of enrollees in Managed Care option have both Part A and Part B

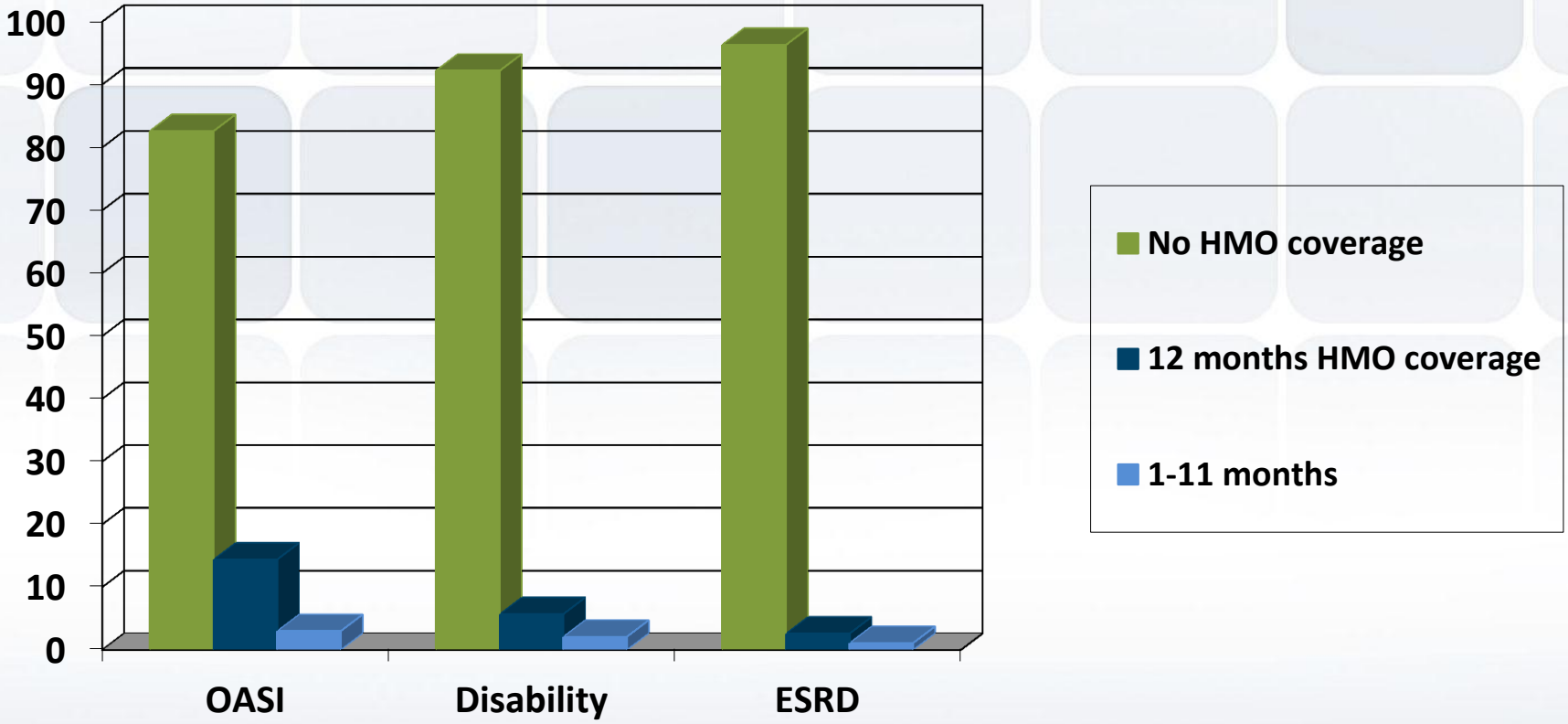
Monthly HMO indicators in the Denominator file

- Indicators do not distinguish between individuals in the FFS system and those not eligible for Medicare benefits.
  - 0 –Not in managed care
  - C–Risk managed care
  - 1–Non-lock-in (cost managed care program)
  - 4—FFS bene in demonstration program; CMS to process claims (new; about 174,000 people (.38%) in 2006 have this value)

Monthly HMO indicators

- **Examples of actual cases:**
- 000000000000 (never in MCO)
- 111111111111 (12 months non-lock-in)
- 00000CC00000 (months 6 & 7 in risk MCO)
- 00CCC0000000 (months 3-5 in risk MCO)
- 000000000000 (never in MCO)
- 111111111111 (whole year in cost MCO)
- CCCCCCCCCCCC (12 months in risk MCO)
- 00000CCCCCCC (months 6-12 in risk MCO)
- 000000000000 (never in MCO)

HMO coverage, 2001



Cost MCOs

- Cost MCOs are a hybrid product created by CMS that is labeled as ‘CMS to process provider claims’
- This label is somewhat misleading.
- For cost MCOs, CMS processes hospital, SNF and Outpatient claims and ONLY A FEW SELECTED types of Carrier claims (some transfusions, PT, etc.)
- Cost MCOs are not found in every market

What you can't tell from the managed care indicators

- Whether someone disenrolled from an MCO by choice or because their area was dropped by plans
- Switching among Medicare HMOs
- Disenrolling from MC due to death (after death all HMO indicators are set to non-HMO)
- Not in an HMO because their Medicare benefits have not yet started or because they have not elected to pay Part B premiums

In the Beneficiary Summary file, we also now can know on a monthly basis:

- H: managed care organization other than regional PPO
- R: Regional PPO
- S: PDP (prescription drug plan)
- N: Not Part D enrolled
- E: Employee-sponsored (beginning in 2007)

For Part D Benefits

- We also can know on a monthly basis:
  - Premium subsidy and copayments
  - Employer subsidy
- But do not know:
  - Specific plans
  - Exact formulary (prior to 2010)

Using a Denominator Record for Defining Populations

- Medicare Status Code=OASI (codes 10 and 11) is a good way to identify elderly
- Most of the time, we would like to limit our studies to 'Persons Likely to Have Complete Claims' this is typically defined as:
  - Both Part A and Part B coverage
  - No managed care enrollment
  - Often, we will require that these conditions be met for some period of time—1 year before to 1 year after hospitalization, etc.
  - If you want to consider Part D coverage too, recognize that this is a somewhat non-random subset…

Using the Denominator Record for Defining Populations

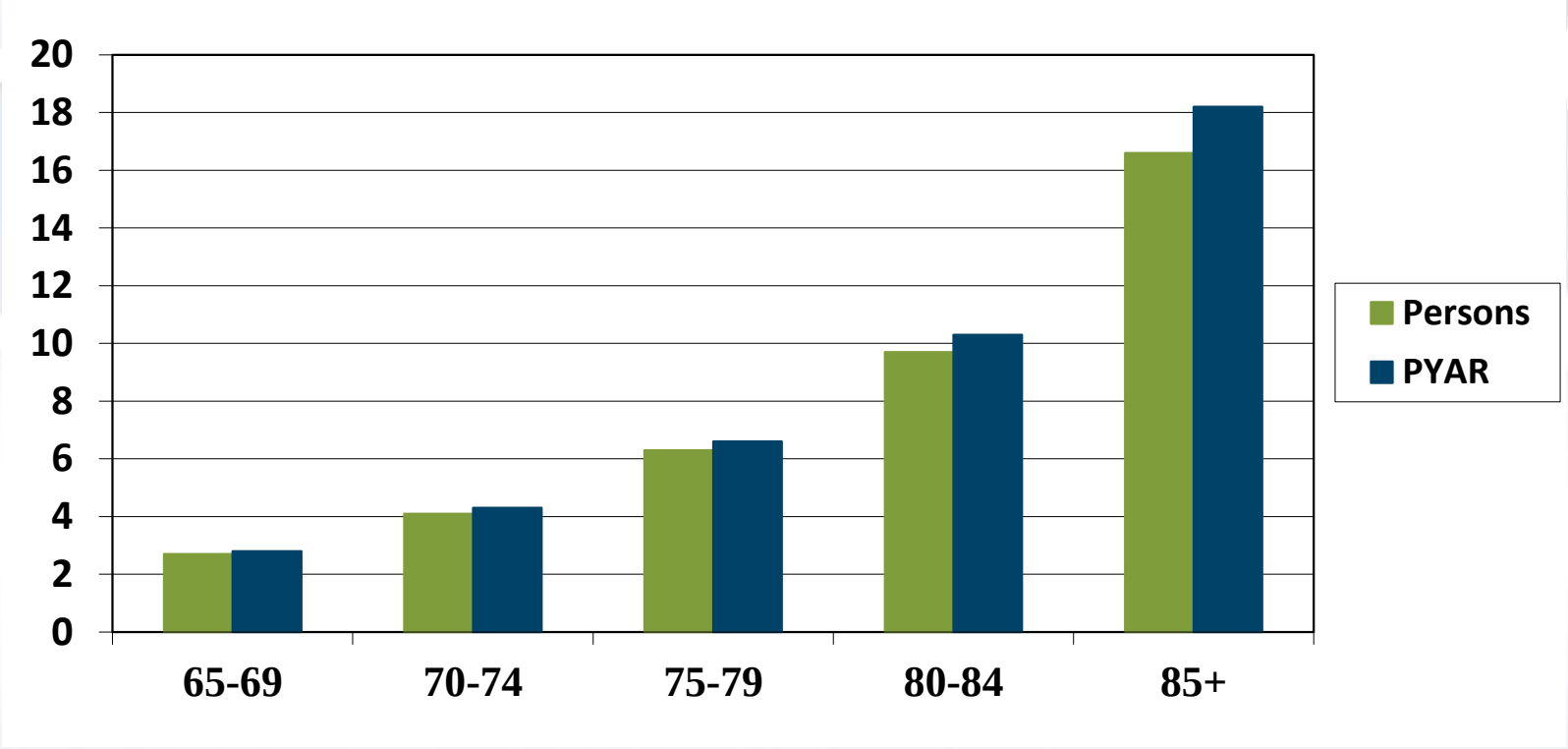
- Medicare Status Code=OASI (codes 10 and 11) is a good way to identify elderly
- Our traditional denominator counts all people equally—assumes all are equally included in the denominator
- Person-months at-risk for death can be calculated using monthly counts–months Part A coverage; this acknowledges that we do not observe all people in the population for the same amount of time
- With the monthly indicators, this approach can be used for mortality, hospitalizations, etc.

At Risk Population—for mortality

Agegp	Sex	Black	PYAR	Number
65-69	1	0	220881.8	241641
65-69	1	1	18127.4	19916
70-74	1	0	171161.3	176260
70-74	1	1	13341.3	13853
75-79	1	0	118944.4	123715
75-79	1	1	8504.1	8959
80-84	1	0	4884.8	75237

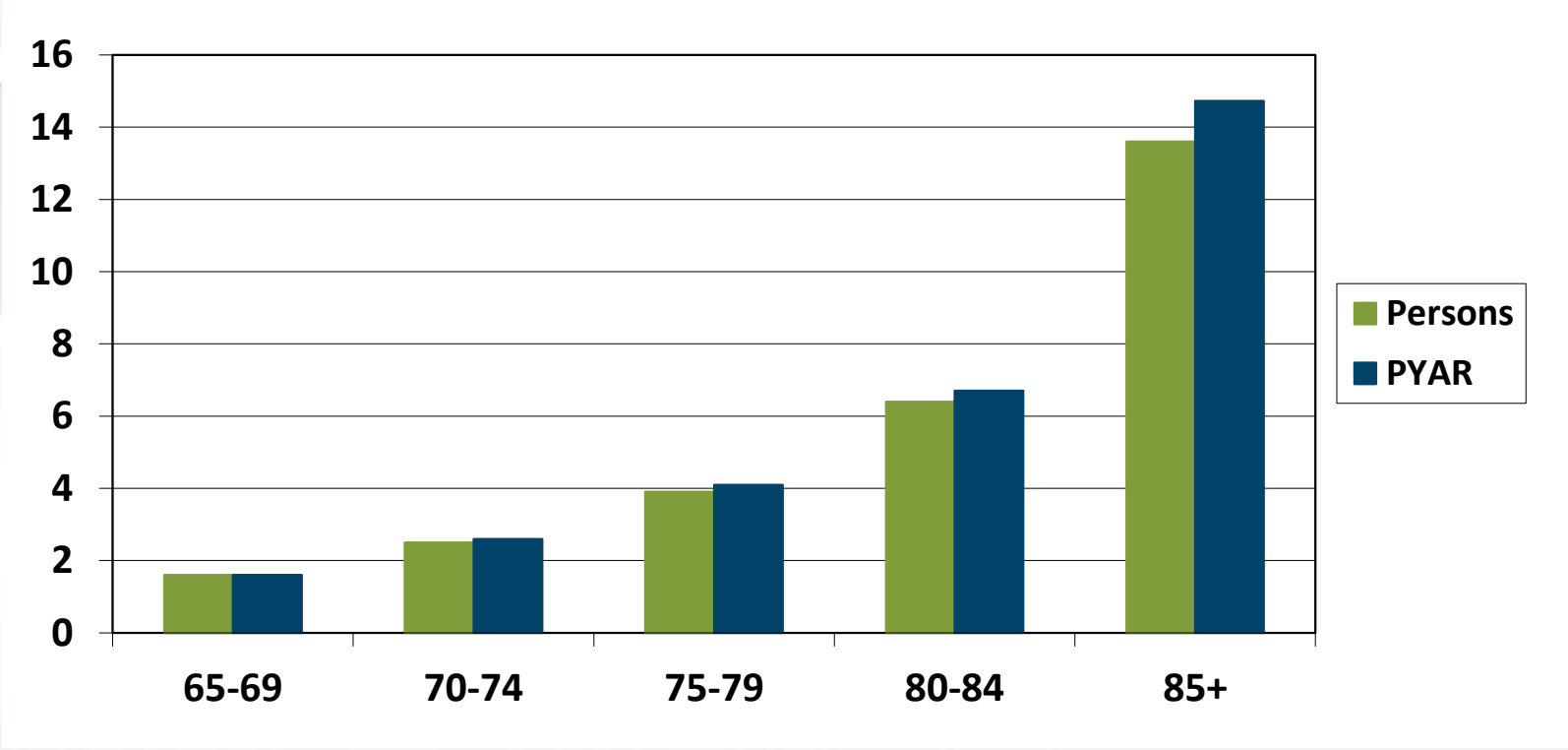
Mortality Rate for Men

PER 100 PERSONS OR PYAR AT RISK

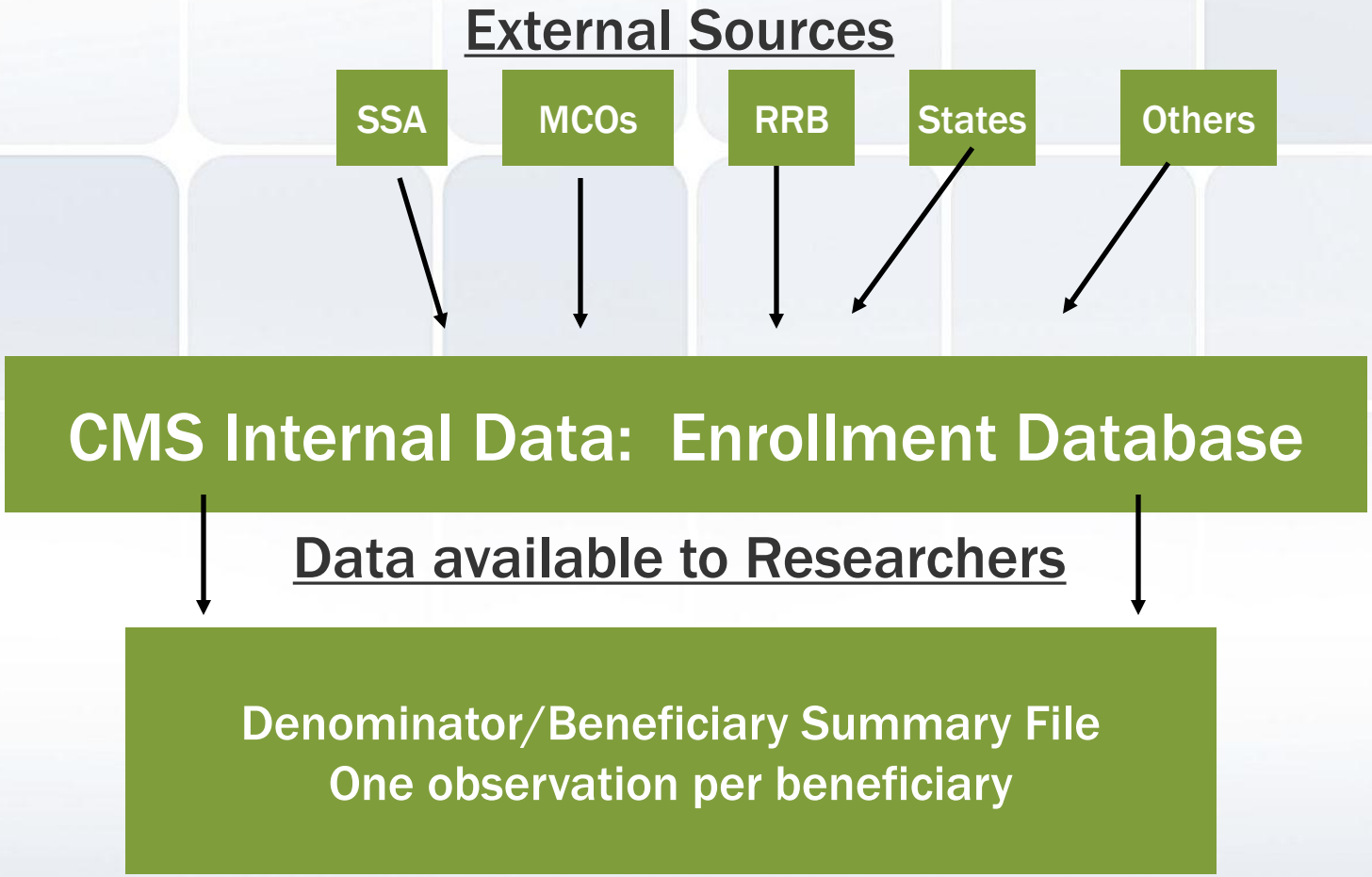


Mortality Rate for Women

(PER 100 PERSONS OR PYAR AT RISK)



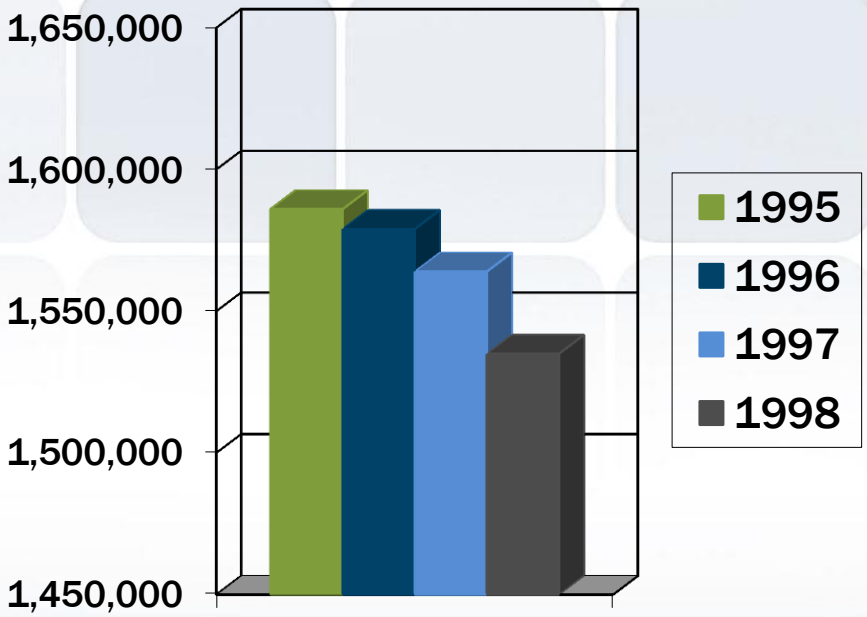
Organization of CMS' Enrollment/eligibility Data



When do you need to order a denominator record?

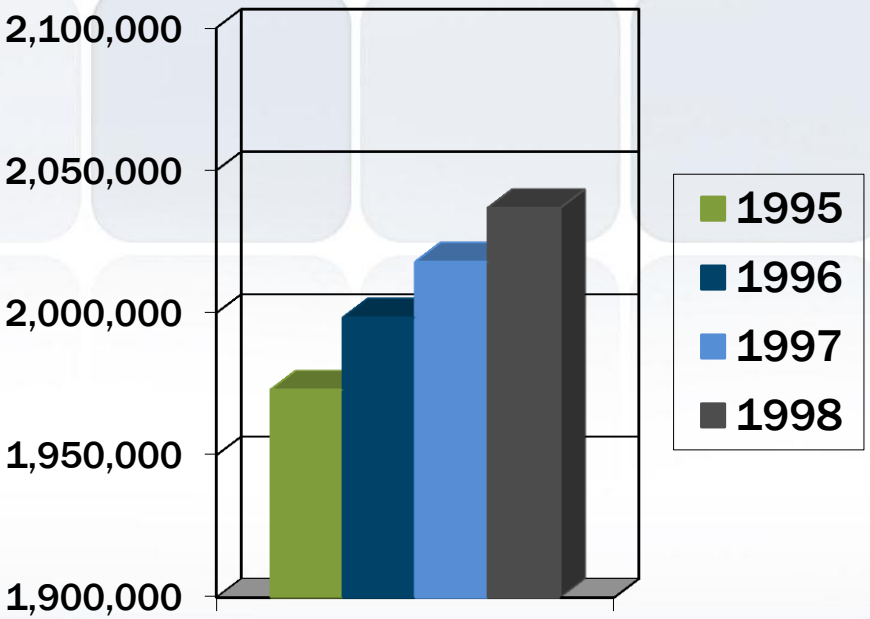
- A denominator record is *always* useful
- A denominator *must* be used when:
  - following a cohort over time (more than 30 days)
    - » To track enrollment/disenrollment and stability of benefit options
  - combining data from an outside source with CMS data
    - » How else can you differentiate no use from not linked?
  - tracking beneficiaries across Part A covered services to Part B covered services
    - » No use vs. didn't elect Part B coverage?
  - Studying beneficiaries who receive benefits through the ESRD or disability programs
    - » People can disenroll from those programs

Example of why a denominator record is always useful



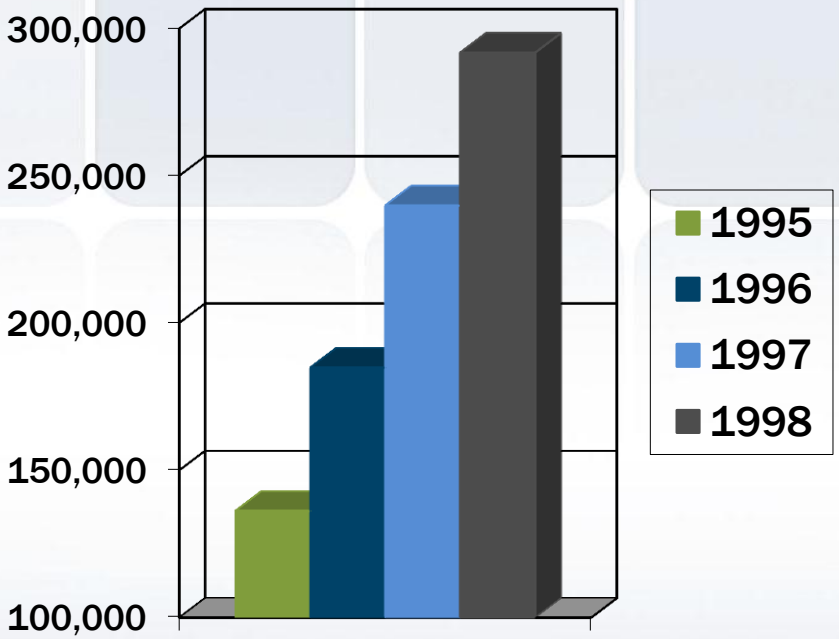
- We have the 5% Carrier Files for 1995-1998. The number of individuals in the file is decreasing
- Number of Medicare recipients SHOULD be increasing
- Physician office visits are unlikely to decrease due to changing practice patterns (vs. hospitalizations)
- Is there a problem with the 5% file?

No, it is fine



- The denominator for the 5% file is increasing, as expected.

But managed care enrollment is increasing faster



- And, at present, Medicare managed care organizations do not submit information on Part B covered services
- So, those individuals will not show up in the Carrier files

Proposed changes to the denominator file

- Indicator of whether information represents beneficiary's physical location or mailing address
- Primary payer code
  - Do they have a primary insurance other than Medicare?
  - What sort of payer is it?