Overcoming Potential Pitfalls in the Use of Medicare Data for Epidemiologic Research

ELLIOTT S. FISHER, MD, MPH, JOHN A. BARON, MD, MSc, DAVID J. MALENKA, MD, JANE BARRETT, MSc, AND THOMAS A. BUBOLZ, PhD

Abstract: We used Medicare data bases and US Census data to address two questions critical to the use of Medicare files for epidemiologic research. First, we examined the degree to which the population enrolled in the Medicare program is similar to the elderly resident population of the United States, as estimated by the US Census. We found small differences in the total population estimates but substantial differences by age and race. Second, we found that among Medicare enrollees, physician claims identify a small proportion of hip fracture cases which are not documented in the hospital discharge files. This proportion varies by age, region, and state within the United States. Calculation of rates based on Medicare hospital discharge data, and probably other hospital discharge data sets as well, must take these limitations into account. Use of all available Medicare data files can overcome these limitations. (*Am J Public Health* 1990; 80:1487–1490.)

Introduction

Administrative health care data bases, such as those maintained by the Health Care Financing Administration (HCFA) for the Medicare program, provide an increasingly accessible and widely used source of data for health care research. Medicare files have been used to study disease incidence and risk factors,[1–3] to document regional variations in treatment patterns,[4,5] to study hospital-associated mortality rates and quality of care,[6] and to evaluate the outcomes of specific surgical and medical treatments.[7,8]

Most concern about the validity of research findings based on the Medicare files has focused on issues related to the accuracy of diagnostic coding.[9–11] In this paper we highlight several other potential limitations of the Medicare data relevant to epidemiologic research by addressing two questions. First, can one assume that the population captured in Medicare files is the same as the elderly population of the United States? Second, do the Medicare hospital claims data identify all discharge records for Medicare enrollees who were hospitalized for a specific disease? The answers to these questions highlight how the careful use of Medicare data can overcome these limitations.

Methods

Population Estimates

We used US Census Bureau survey data and Medicare data files to estimate the age, sex, and race-specific counts of United States residents and Medicare enrollees who were over the age of 65 and were alive on July 1, 1985. Census estimates were taken from the Current Population Estimates.[12] Estimates for the Medicare eligible population were determined by using the 5 percent National HISKEW (live) file maintained by the Health Care Financing Administration. This file is a random sample of the Medicare population, based on the terminal two digits of the social security number and records the birth date, state of residence, race, gender, and Medicare eligibility status of beneficiaries who have been eligible for Medicare at some point and are not known to have died. (Deaths are recorded in a different file.) We excluded Medicare eligibles listed as residing outside the 50 States. We then determined age, sex, and race-specific counts of individuals in the HISKEW files generated on 12/31/84 and on 12/31/85. We averaged these counts and multiplied by 20 to estimate the Medicare population alive on July 1, 1985.

Hip Fracture Case Ascertainment

We searched the 5 percent National MEDPAR (hospital discharge) files for all discharges in calendar year 1985 with a discharge diagnosis of hip fracture (ICD-9-CM code 820.0–820.9) and the 5 percent Medicare Part B (physician and other provider) files for all claims specifically documenting the treatment of an acute hip fracture during calendar year 1985 (Current Procedure Terminology codes in the range 27230 to 27248). We developed a detailed algorithm to define acute hip fractures based on all claims submitted during the calendar year for potential cases, relying on both diagnosis and procedure codes and the dates of the related claims.[13] To be included, cases had to have at least one hospital claim specifying the diagnosis, or two Part B claims supporting the diagnosis of hip fracture. We excluded potential cases for whom the only evidence of treatment for the fracture was provided by a single Part B claim (3 percent of candidate cases). This decision was based on extensive review of all claims for a sample of these potential cases. Hip fracture cases were excluded if there was evidence that the treatment was provided for an earlier fracture or that the fracture resulted from metastatic cancer. We applied this algorithm to the data obtained from the 1985 5 percent national sample of Part A and Part B claims.

Results

Population Estimates

We compare the US Census estimate of the elderly population with our estimate of the Medicare population in Table 1. Overall, the Medicare population, as estimated from the HISKEW file, represents approximately 96 percent of the US Census estimate of the elderly population, but the proportion enrolled in Medicare varies sharply by both age and race. Among Whites, only 90 percent of the estimated elderly population between the ages of 65 and 69 is enrolled in Medicare and the proportion enrolled increases with age. Among Blacks, only 79 percent of the estimated population

From the Department of Medicine, Dartmouth Medical School (Fisher, Baron, Malenka); Department of Community and Family Medicine at Dartmouth (all authors); and the Department of Veterans Affairs Medical Center, White River Junction, Vermont (Fisher). Address reprint requests to Elliott S. Fisher, MD, MPH, Department of Community and Family Medicine, Dartmouth Medical School, Hanover, NH 03756. This paper, submitted to the Journal May 9, 1990, was revised and accepted for publication August 12, 1990.

© 1990 American Journal of Public Health 0090-0036/90$1.50

TABLE 1—Comparison of Population Estimates based on Medicare Enrollment File (HISKEW) and US Current Population Survey (counts in thousands)

	Male				Female				All			
Race/Age	HSK	CNS	H/C	C-H	HSK	CNS	H/C	C-H	HSK	CNS	H/C	C-H
White												
65–69	3425	3821	0.90	396	4196	4615	0.91	419	7620	8436	0.90	816
70–74	2705	2899	0.93	194	3674	3914	0.94	240	6380	6813	0.94	433
75–79	1797	1924	0.93	127	2880	3044	0.95	164	4677	4968	0.94	291
80–84	1021	1051	0.97	30	1998	2001	1.00	3	3019	3052	0.99	33
80+	674	695	0.97	21	1760	1779	0.99	19	2434	2474	0.98	40
Total	9623	10390	0.93	767	14508	15313	0.95	805	24131	25743	0.94	1612
Black												
65–69	287	361	0.79	74	374	472	0.79	98	661	833	0.79	172
70–74	230	258	0.89	28	331	373	0.89	42	561	631	0.89	70
75–79	152	172	0.89	20	250	270	0.93	20	403	442	0.91	39
80–84	87	84	1.04	−3	164	149	1.10	−15	251	233	1.08	−18
85+	65	65	0.99	0	146	140	1.04	−6	210	205	1.03	5
Total	821	940	0.87	119	1265	1404	0.90	139	2086	2344	0.89	258
Other Race												
65–69	65	72	0.91	7	79	89	0.89	10	144	161	0.90	17
70–74	53	56	0.95	3	59	66	0.90	7	113	122	0.92	9
75–79	36	39	0.91	3	31	46	0.67	15	66	85	0.78	19
80–84	20	19	1.05	−1	18	26	0.70	8	38	45	0.84	7
85+	11	13	0.81	2	14	19	0.75	5	25	32	0.77	7
Total	185	199	0.93	14	201	246	0.82	45	386	445	0.87	59
Unknown/missing race												
65–69	118	0	—	−118	155	0	—	−155	273	0	—	−273
70–74	97	0	—	−97	130	0	—	−130	227	0	—	−227
75–79	62	0	—	−62	106	0	—	−106	168	0	—	−168
80–84	28	0	—	−28	55	0	—	−55	83	0	—	−83
85+	16	0	—	−16	57	0	—	−57	74	0	—	−74
Total	322	0	—	−322	503	0	—	−503	825	0	—	−825
All Races												
65–69	3895	4254	0.92	359	4803	5176	0.93	373	8698	9430	0.92	732
70–74	3086	3213	0.96	127	4194	4353	0.96	159	7281	7566	0.96	285
75–79	2047	2135	0.96	88	3268	3360	0.97	92	5315	5495	0.97	180
80–84	1156	1154	1.00	−2	2235	2176	1.03	−59	3391	3330	1.02	−61
85+	766	773	0.99	7	1978	1938	1.02	−40	2743	2711	1.01	−32
Total	10951	11529	0.95	578	16478	17003	0.97	485	27428	28532	0.96	1104

Notes: Estimates of Medicare enrolled population calculated by taking average of live counts for 5% national sample for 12/31/84 and 12/31/85 in each relevant cell and multiplying by 20. Persons recorded as on the live HISKEW file but who were not residents of the 50 states were excluded. Census estimates were for 7/1/85, and were obtained from the Current Population Survey.[12]

HSK = Estimate of population from Medicare's HISKEW file.
CNS = Estimate of population from Current Population Survey.
H/C = Ratio of HISKEW estimate to CPS estimate.
C-H = Census estimate minus HISKEW estimate

aged 65 to 70 is enrolled in Medicare. Among Blacks, the proportion enrolled increases with age and exceeds the Census estimates among those over 80 years old. No notable differences between the two genders in the age-specific proportions enrolled in Medicare are apparent for Whites or Blacks. However, among those whose race is designated as "other," different patterns are apparent for males and females. Finally, for 3.1 percent of the individuals in the HISKEW file, race is designated as unknown, while the Census assigns a race designation to all individuals.

Case Ascertainment from Hospital Claims

Table 2 presents data on the proportion of hip fracture cases identified from the Part B (physician and other provider) files by state for 1985. Overall, 9.1 percent of hip fracture cases were identified from the Part B claims alone, and the proportion of cases so identified ranged from none in several states to a high of 32.8 percent in Virginia. The proportion of cases identified only through the Part B claims decreased significantly with age from 10.6 percent for those ages 65 to 69 to 8.4 percent for those age 85 and over (Chi square, 4df, 12.9, p = .01). However, no significant differences were found in the proportion of cases identified through the Part B files by gender or race.

Discussion

It is often assumed that the Medicare-eligible population corresponds closely to the elderly population of the United States and that studies of Medicare's hospital discharges will provide accurate estimates of hospital utilization among the elderly. One recent study has estimated hip fracture incidence rates using data from the US Census to estimate age/race/gender-specific denominators, while using Medicare hospital discharge data to calculate age, race, and gender-specific numerators for these rates.[3] Our data suggest that this approach must be taken with considerable caution.

We found that the Medicare population over the age of 65 differs in several important ways from the US Census estimates of the elderly population as a whole. The fact that only 95 to 96 percent of US elderly residents are enrolled in Medicare has been previously reported,[14] but breakdowns by age, gender, or race were not provided. Our data demonstrate that Medicare coverage varies substantially across these groups. Although some of the differences between Census and Medicare estimates of the population are trivial, others

TABLE 2—Proportion of Hip Fracture Cases Identified through the Medicare Part B Claims only, by State, based on 5% National Sample, Calendar Year 1985

State/Region	N Cases	% B Only
NORTHEAST		
Connecticut	128	0.0
Maine	46	4.3
Massachusetts	249	13.3
New Hampshire	38	7.9
New Jersey	276	18.1
New York	685	11.1
Pennsylvania	479	7.9
Rhode Island	48	12.5
Vermont	19	0.0
Region Total	1968	10.6
MIDWEST		
Illinois	439	11.4
Indiana	262	7.6
Iowa	156	4.5
Kansas	132	3.0
Michigan	268	9.7
Minnesota	151	4.6
Missouri	254	4.7
Nebraska	63	4.8
North Dakota	33	9.1
Ohio	443	11.5
South Dakota	32	3.1
Wisconsin	208	2.4
Region Total	2441	7.7
SOUTH		
Alabama	141	10.6
Arkansas	109	5.5
Delaware	27	3.7
District of Columbia	9	11.1
Florida	522	11.7
Georgia	186	7.5
Kentucky	159	6.9
Louisiana	99	4.0
Maryland	111	9.9
Mississippi	78	7.7
North Carolina	179	27.4
Oklahoma	146	4.8
South Carolina	68	5.9
Tennessee	195	4.6
Texas	521	7.5
Virginia	131	32.8
West Virginia	78	6.4
Region Total	2759	10.4
WEST		
Alaska	3	0.0
Arizona	107	5.6
California	821	8.6
Colorado	107	9.3
Hawaii	10	0.0
Idaho	33	3.0
Montana	29	3.4
Nevada	24	0.0
New Mexico	43	2.3
Oregon	123	5.7
Utah	27	0.0
Washington	163	6.7
Wyoming	15	6.7
Region Total	1505	7.2

are not. Consideration of sampling variations would further complicate interpretation of these differences. However, our analyses underline the obvious but important point that Medicare data provide information about the Medicare population; findings of studies that rely on Medicare data should be generalized to the elderly population as a whole with caution.

A full exploration of the causes of these differences is beyond the scope of this paper but would include issues related to criteria for Social Security eligibility, delayed enrollment for those still employed after age 65, and decisions by eligible persons about whether to enroll in either or both of the Medicare Part A and Part B programs. That the Current Population Estimates data from the US Census bureau are estimates subject to both sampling and measurement error may also play a role.

Equally important, coding conventions for race used by the Census and by Medicare differ. While the Census assigns a racial designation to all residents, approximately 3 percent of Medicare enrollees remain of unknown race. Given the White predominance in the elderly population and the possibility that substantial numbers of Whites are included in the group designated unknown, it would be inappropriate to group those of unknown race with either of the other two minority groups.

Finally, we must note that we cannot explain the relative excess of Medicare eligibles compared to Census estimates for the older age groups in several racial categories. Although Census undercounts in the oldest population are conceivable, this could also be explained by the failure to record a small number of deaths in the HISKEW file, a deficiency that would be magnified in the oldest age categories. We have found some indication that this might be the case. The 1985 5 percent national HISKEW file included 368 individuals over the age of 110, resulting in an estimated 7,360 individuals over this age in the population as a whole—an implausible figure. Nevertheless, given the close similarity of our overall estimates of the Medicare population based on HISKEW to HCFA's estimate of the proportion of the US population enrolled in Medicare (95 percent[14]), few deaths overall are apparently missed by these files and analytic difficulties caused by this problem would be serious only among the oldest old.

While it is obvious that outpatient claims would be required to analyze incidence rates for fractures that can be treated in an outpatient setting, we found that Medicare's hospital discharge files do not capture all treatment for hip fractures, a condition virtually always treated on an inpatient basis. Overall, physician claims were required to identify 9.1 percent of hip fracture cases; almost all of these physician claims were coded with the place of service "inpatient." For the majority of these cases, we found associated hospital claims with evidence of treatment for orthopedic conditions related to the hip, but no diagnosis of hip fracture. For the remainder, we must conclude that the hospital claims were either lost in the data transmission process or severely miscoded. It is reassuring that only small and insubstantial differences in fracture detection rates were found by age, race or gender, but the variation by state of residence might pose problems for analyses of geographic differences in incidence rates across states. We should also note that gaps in the physician claims are possible—Connecticut Part B claims were missing from the 1985 Medicare Part B 5 percent national file. A second caveat is that some Medicare enrollees may receive care in Veterans' Administration hospitals. We found that approximately 4 percent of New England hip fracture cases among males (and less than 1 percent among females) were documented in VA files.[13]

Our data illustrate two specific limitations beyond the accuracy of diagnostic coding that must be considered when using or reporting findings of studies that rely on Medicare data. First, the Medicare population cannot be assumed to correspond exactly with the resident US elderly population. Second, gaps in the Medicare hospital data must be acknowl-

edged. These potential limitations could also apply to any analysis where different data sources are used for the numerator and denominator, or where hospital discharge data sets are not complete prior to release as public use tapes.

Most important, however, our data underline the feasibility of overcoming these limitations through the careful linkage of data from the multiple files available through the Medicare program. One of the cardinal rules in calculating rates is to ensure that the criteria used to determine eligibility for inclusion in the numerator should be as similar as possible to those used for assignment to the denominator. The Medicare enrollment file (HISKEW) can be used not only to define precisely denominators by age, race, sex, and mailing address (zip code), but can also be used to define identical inclusion criteria for both the numerator and denominator of rates. Finally, physician claims obtained from the Part B files and hospital claims from the Veterans' Administration can be merged with the Medicare hospital claims files to reduce the likelihood of missing cases in the study population of interest. When the limitations of administrative health care data bases are recognized and appropriate care exercised in their use, they can provide a valuable tool for epidemiologic research.

REFERENCES

1. Silverman SL, Madison RE: Decreased incidence of hip fracture in Hispanics, Asians and Blacks: California hospital discharge data. Am J Public Health 1988; 78:1482–1483.
2. Farmer ME, White LR, Brody JA, Bailey KT: Race and sex differences in hip fracture incidence. Am J Public Health 1984; 74:1374–1380.
3. Kellie SE, Brody JA: Sex-specific and race-specific hip fracture rates. Am J Public Health 1990; 80:326–328.
4. McPherson K, Wennberg JE, Hovind OB, Clifford P: Small-area variations in the use of common surgical procedures: An international comparison of New England, England, and Norway. N Engl J Med 1982; 307:1310–1314.
5. Wennberg JE, Freeman JL, Shelton RM, Bubolz TA: Hospital use and mortality among Medicare beneficiaries in Boston and New Haven. N Engl J Med 1989; 321:1168–1173.
6. Health Care Financing Administration: Medicare Hospital Mortality Information 1987. Washington, DC: Govt Printing Office, 1988.
7. Roper WL, Winkenwerder W, Hackbarth GM, Krakauer H: Effectiveness in health care. An initiative to evaluate and improve medical practice. N Engl J Med 1988; 319:1197–1202.
8. Roos NP, Wennberg JE, Malenka DJ, *et al*: Mortality and reoperation after open and transurethral resection of the prostate for benign prostatic hyperplasia. N Engl J Med 1989; 320:1120–1124.
9. Jencks SF, Williams DK, Kay TL: Assessing hospital-associated deaths from discharge data. The role of length of stay and comorbidities. JAMA 1988; 260:2240–2246.
10. Greenfield S, Aronow HU, Elashoff RM, Watanabe D: Flaws in mortality data. The hazards of ignoring comorbid disease. JAMA 1988; 260:2253–2255.
11. Iezzoni LI, Burnside S, Sickles L, Moskowitz MA, Sawitz E, Levine PA: Coding of acute myocardial infarction: clinical and policy implications. Ann Intern Med 1988; 109:745–751.
12. US Bureau of the Census, Department of Commerce, Current Population Reports, Series P-25, No. 985, Table 3. Estimates of the population of the US by age, race and sex, 1980–1985. Washington, DC: Govt Printing Office, 1986.
13. Fisher ES, Baron JA, Malenka DJ, Barrett JA, Kniffin WD, Whaley FS, Bubolz TA: Hip fracture incidence and mortality in New England. Epidemiology (In press).
14. Hatten J: Medicare's common denominator: The covered population. Health Care Financ Rev 1980; 2:53–63.

Future APHA Meeting Dates/Site

1991 NOVEMBER 10–14 ATLANTA, GEORGIA
1992 NOVEMBER 8–12 WASHINGTON, DC
1993 OCTOBER 24–28 SAN FRANCISCO, CALIFORNIA
1994 OCTOBER 30–NOVEMBER 3 WASHINGTON, DC