

Matching MCBS and Medicare Data: The Best of Both Worlds

Franklin J. Eppig, J.D., and George S. Chulis, Ph.D.

Survey reports from the Medicare Current Beneficiary Survey (MCBS) were matched to Medicare administrative files to create the 1992 MCBS Cost and Use file. This file improves on previous MCBS Access-to-Care user files by representing the entire (ever enrolled) Medicare population and including services not covered by Medicare such as outpatient prescription drugs and long-term facility care. The matching and reconciliation process improved the accuracy and completeness of health care use and cost. For example, Medicare billing data corrected 22 percent of survey reports that did not record Medicare as a payer and 39 percent in which the amount was missing.

BACKGROUND

The MCBS is an ongoing household panel survey of approximately 12,000 elderly and disabled persons eligible for Medicare benefits.¹ Field work for the MCBS began in September 1991. To date, five MCBS Access-to-Care Public Use Files (PUFs)—1991 through 1994—have been produced and made available to the public. The Access-to-Care PUFs link survey data on access to and satisfaction with health care, supplementary health insurance, and health and disability status, which are typically collected in the fall round each

year, to Medicare billing data that cover the entire calendar year.

These PUFs have been used extensively to analyze a variety of issues, including: access to health care (Physician Payment Review Commission, 1996; Rosenbach, Adamache, and Khandker, 1995); satisfaction with health care (Adler, 1995); premium payments for supplementary health insurance (Chulis, Eppig, and Poisal, 1995); the relationship between supplementary health insurance and Medicare spending (Chulis et al., 1993); risk adjusting per capita payments to Medicare health maintenance organizations (HMOs) (Gruenberg, Kaganova, and Hornbrook, 1996); examining favorable HMO selection (Rodgers and Smith, 1996); and the characteristics of users of home health services (Mauser and Miller, 1994).

There are, however, some significant analytic limitations to the MCBS Access-to-Care PUFs. One limitation relates to the population covered. The Access-to-Care files represent the "always enrolled," that is, elderly and disabled Medicare beneficiaries entitled to Medicare for the entire calendar year. This enrollment concept excludes persons who come on the Medicare rolls during the year. More significantly, it excludes most persons who died during the year. Persons in this group have medical expenses that are considerably higher on average than surviving beneficiaries (Lubitz and Riley, 1993).

Another limitation of the Access-to-Care PUFs is that they do not contain survey-

The authors are with the Office of the Actuary, Health Care Financing Administration (HCFA). The opinions expressed are those of the authors and do not necessarily reflect those of HCFA.

¹ See Adler (1994) for a full description of the MCBS.

reported use of health services and costs. The files do include use and payments for Medicare covered services from Medicare billing records. However, Medicare covers less than one-half of total health care expenditures for the elderly (Waldo et al., 1989). Two of the more financially significant health care services not covered by Medicare, and therefore not included in the Access-to-Care files, are outpatient prescription drugs and long-term facility care.

The 1992 MCBS Cost and Use PUF is designed to create a more complete user file, one that uses an "ever enrolled" population concept and that includes all survey-reported use and costs. The "ever enrolled" population includes use and costs for all Medicare beneficiaries in the program for any part of 1992, including those who joined the program during the year and those who died during the year. The Cost and Use file also includes survey reports for services not included in Medicare central billing files, including prescription drugs, long-term facility care, and Medicare services provided by HMOs. In addition, for Medicare covered services, the completeness and accuracy of services used, payments made, and sources of payment has been improved by an extensive operation to match and reconcile survey reports and Medicare bills. This article describes the methods used and the results from the matching and reconciliation process used to create the 1992 MCBS Cost and Use file.

MATCHING SURVEY AND ADMINISTRATIVE REPORTS

There has been a continuing emphasis in government-sponsored research to find better ways to use government administrative records to verify and augment information reported on surveys (Okner, 1974; Jabine and Scheuren, 1984). The

advantages of linking survey reports to administrative records include verifying the accuracy of survey reports, adding data that was not (or could not be) obtained in the survey, and reducing the reporting burden on respondents. In health surveys, in particular, it has long been recognized that respondents may not be the best source of information on payments for health care services, particularly if health insurance companies are making payments directly to providers on the respondent's behalf (Cohen and Carlson, 1994).

In the case of Medicare inpatient hospital services, for example, a lump-sum payment is made directly to a hospital based on the patient's diagnosis-related group (DRG). A beneficiary reporting an inpatient hospital stay rarely if ever would know the amount the Medicare program paid the hospital. For other types of Medicare services, such as physician visits, the beneficiary is notified of Medicare program payments in an Explanation of Medicare Benefits (EOMB) form. In the survey interview setting, however, the typical respondent is usually able to recall the amount that they paid out-of-pocket for a health service, but is less clear on the amount that Medicare, a private supplementary insurer, or other third party paid on their behalf.²

Table 1 shows item non-response rates for selected variables in the MCBS in 1992. It makes clear that demographic and socio-economic questions are much better reported than charge and payment amounts. Since accurate and complete information on use of services, payments, and sources of payment are the primary objectives of the MCBS, item non-response rates of 25 percent and higher for charge and

² To eliminate the need for the respondent to search his or her memory or be forced to guess third-party payment amounts, the MCBS interview relies heavily on information from Medicare and private insurance statements.

Table 1
MCBS Non-Response Rates for
Selected Variables: 1992

Variable	Percent Missing
Race	0.1
Ethnicity	0.3
Education	3.4
Marital Status	0.2
Gender	0.0
Age	0.1
Total Charge for Health Event	26.0
Total Payment for Health Event	30.7

NOTE: MCBS is Medicare Current Beneficiary Survey.

SOURCE: 1992 MCBS Survey Report file.

payment questions are not acceptable. Anticipating this difficulty, it was understood from the planning stages of the MCBS that survey-reported dollar amounts would have to be verified and augmented using Medicare billing data.

Other national health surveys, such as the 1987 National Medical Expenditure Survey (NMES), also relied on matching survey reports to provider records (Cohen and Carlson, 1994). However, due to the expense involved in matching and reconciliation, the NMES provider follow-back to verify survey reports was limited to a 25-percent sample. (The sample was judiciously constructed to over-represent events and persons which would give good returns to provider follow-back. The strategy included disproportionately high shares of expensive hospital claims and Medicaid enrollees, who are particularly poor survey respondents.) A limitation of the NMES provider follow-back survey was that it relied on the survey reports to identify the providers to be contacted. While this approach would undoubtedly improve accuracy for survey-reported health events, it was more limited in correcting for events that were omitted. That is, if a respondent did not report a health event, and that provider was not reported for another event by that person during the interview, the missed event would never have a chance to be detected in the provider follow-back survey.

Under-reporting of health events is a serious problem in health surveys. In general, the farther removed in time a health event is from the interview date and the less salient or significant the health event is in the person's life, the higher the likelihood it will not be reported in the interview.³ As a rule, respondents remember and report inpatient hospitalizations better than doctor visits, and doctor visits in the last 2 weeks better than those occurring 2 months ago. The recall period for MCBS interviews is usually about 4 months. Unfortunately, this problem cannot simply be solved by more frequent interviews and shorter recall periods. In addition to the considerable extra expense that would be involved in interviewing more frequently, there is evidence from an earlier national health panel survey that expected gains in recall by more frequent interviewing and shorter reference periods could be offset by negative "conditioning effects" due to increased reporting burdens on respondents (Cohen and Burt, 1985).

For these reasons, conducting a match that will detect both survey under-reporting and reporting inaccuracy is clearly preferable. Fortunately, virtually complete billing records of services used and payments made under Medicare fee-for-service transactions are kept in HCFA files.⁴ The MCBS—a HCFA-sponsored survey—was uniquely positioned to do a match of survey reports to billing records. Unlike other health surveys, the MCBS was designed from the start to be a full partner with Medicare administrative records. Survey reports have been joined to bill records to form a more complete and

³ An excellent summary of the literature on recall periods and reporting accuracy can be found in Cohen and Burt (1985).

⁴ The primary group for whom there are missing or incomplete central office bill files, and therefore could not be included in the match, are persons enrolled in Medicare managed care plans. In 1992, this represented 6 percent of Medicare enrollees. An estimated 97 percent of Medicare claims are posted to HCFA central billing records within 1 year (Eppig and Edwards, 1996).

accurate file than would be possible using either source alone.

CRITERIA FOR THE MATCH

In terms of survey methods, this is an “exact” match of information for the same person from two different data sources, not a synthetic or “statistical” match which imputes information to an individual based on similarities in key characteristics (Okner, 1974). The unique health insurance claim number (HICN) for each Medicare enrollee is recorded on each of the central office administrative billing records. A sample person’s HICN is known when he or she is selected for the MCBS. Sample persons are asked in the first interview to verify their HICN by showing their Medicare identification card, and their HICN number is then permanently associated with all subsequent survey records. This unique personal identifier that is common to both record sources insures that all survey and administrative billing records for each person can be pulled together prior to the match.⁵

Although administrative records hold out the potential for improving survey reports, previous experience has shown that the matching process is never straightforward, and that it is not wise to simply assume that the administrative data is the “correct” source. Previous efforts to match survey reports to administrative records have shown that both data sources in a match, not just survey reports, invariably have limitations which complicate the matching process and the interpretation of results. As Winn and Walden commented in a review of several methods studies that examined

matches of survey reports to administrative records, “survey researchers should not use administrative record data as a ‘gold standard’ or even ‘gold plated standard’” (Winn and Walden, 1989).⁶ The limitations of administrative records when matched to survey reports are generally not due to poor quality record keeping, but rather stem from differences in the basic purposes for which the records were created. In discussing an analysis of a match of survey reports to Medicare bills, Verbrugge made the same point this way: “Billing systems have motivations quite unrelated to patient care...There is no one to one relationship between visits and bills. To compare them, that relationship has to be constructed...” (Verbrugge, 1989).

Our approach in constructing the match between MCBS survey reports and Medicare central office billing records—dissimilar records collected for different purposes—was that neither source should be considered a “gold standard.”⁷ Each source has its strengths and limitations. For each item of information collected from both sources, decisions were made based on the likelihood that one source would be more accurate or complete than the other source in the context of that particular comparison. The objective was a combined record that embodies the best features of each data source, and that was more accurate and complete than either the survey reports or billing records used alone. In general, Medicare bill records were thought to be the more accurate source for information:

⁵ While the HICN was used to create the Cost and Use file, it does not appear in the user file because this would violate the sample person’s right to privacy.

⁶ Among the studies reviewed by Winn and Walden was a match of hospitalizations reported on the 1987 NMES survey to HCFA’s Medicare Automated Data Retrieval System (MADRS). (Calore and Lim, 1989).

⁷ A discussion of the match methods and some early results were published in Eppig and Edwards (1996).

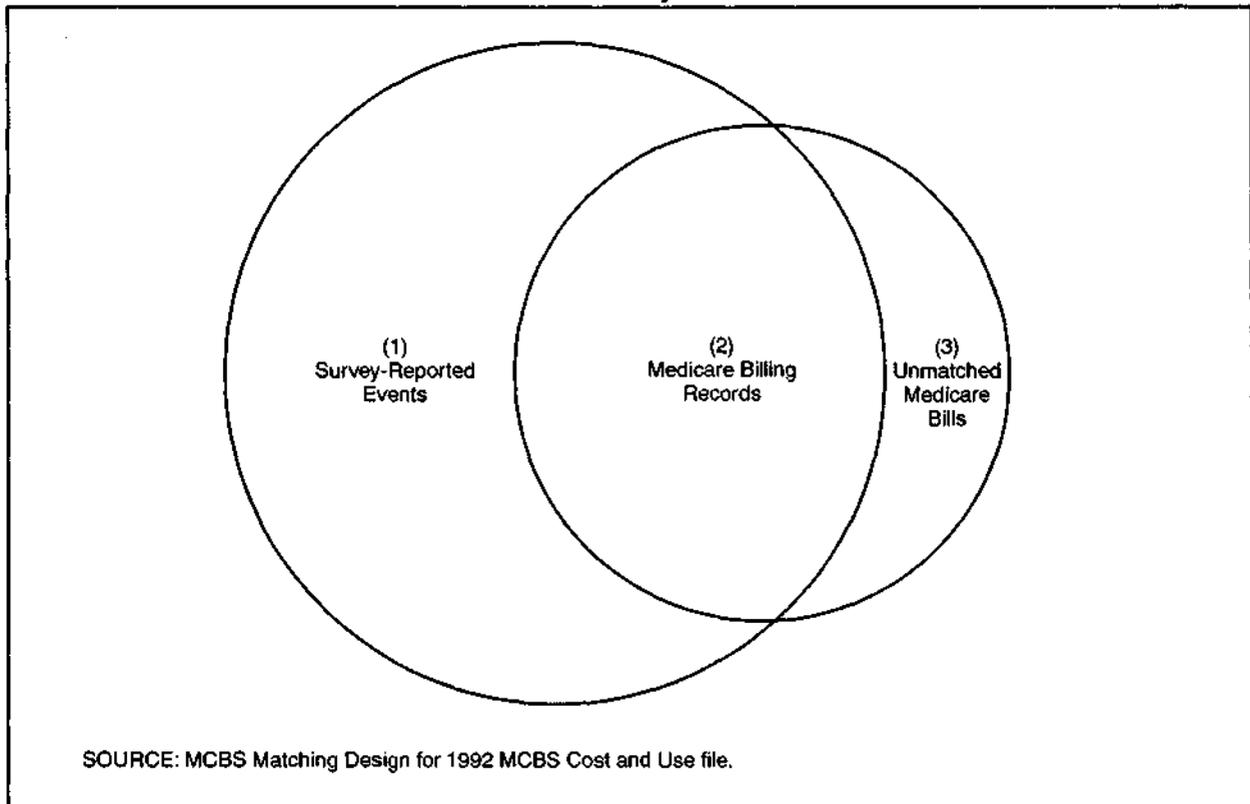
- That a health service occurred. (A billing record showing Medicare payment for a service that was not reported on the survey was considered a memory lapse by the respondent.)
- That Medicare was a payer. (A billing record showing Medicare payment for a service was considered to be more accurate than the respondent not reporting Medicare as a payer.)
- On the amount paid by Medicare. (The Medicare payment amount in the records was considered more accurate than the amount reported by the beneficiary. As noted above, there are good reasons why a beneficiary would not know the DRG payment for inpatient hospital services or the amount paid the physician under the relative value based physician fee schedules.)

In most other situations, and particularly regarding reports of amounts paid out of pocket, the survey reports were given precedence. Figure 1 illustrates three important issues that had to be considered in designing the match of MCBS survey-reported events to Medicare central office billing records.

NARROWING DOWN SURVEY REPORTS

Survey-reported events (labeled "1" in Figure 1) are broader in scope than Medicare billing events because the survey collects information on all health services, including services not covered by Medicare. This means that survey events that are clearly not Medicare covered services, such as prescription drugs, must be eliminated prior to the match. However,

Figure 1
Schematic of Survey-to-Bill Match



this process must be done carefully to avoid removing any survey events that could conceivably match a Medicare billing record had they been left in the match. On the other hand, including types of survey events for which a match is possible but not probable could increase the chance of “false positives,” matches that qualify according to matching criteria, but are not really genuine. After considering these tradeoffs in light of the variables to be used in establishing a match, we decided to use a fairly unrestricted approach. For example, we included all survey reports of dental services in the match even though Medicare rarely pays for dental procedures.⁸ In part, this reflected our desire to match as many survey reports as possible. But we also judged that, given the specific variables used to identify a match, the risk of false positive matches was not large. (One clear implication of this decision to broaden the types of survey reports included in the match is that a large share of survey reports can reasonably be expected not to match a Medicare bill.)

COMPARABILITY OF SURVEY RECORDS AND BILLS

A significant number of events—the expected matches—will appear in both the survey events and Medicare billing records (labeled “2” in Figure 1). However, Medicare billing records often do not record events in the same way that they were reported in the survey. A respondent may report a visit to a physician as a single event on the survey. However, Medicare’s fee-for-service billing records may have recorded separate payments for a physician service, an X-ray, a laboratory test, supplies

⁸ Medicare does not cover routine dental care and only pays for dental procedures when they can be shown to be integrally related to other strictly medical procedures, e.g., tooth extraction as part of jaw surgery.

such as bandages, etc. for that same visit. This means that extensive effort is required to put survey-reported “events” and Medicare billing “events” on the same basis prior to matching. It also raises the philosophical question of which concept of an event is more appropriate, and just how far it is desirable to go in shaping survey reports to look like Medicare events or vice versa. We discuss these issues in more detail later.

UNMATCHED MEDICARE CLAIMS AND SURVEY REPORTS

A final point is brought out by Figure 1. If ordinary presuppositions about survey under-reporting because of memory decay over a 4-month reference period are correct, we would expect there to be unmatched Medicare billing records. These medical events (labeled “3” in Figure 1) represent services paid for by Medicare that were not reported on the survey.

In many matches, the size of the share of matched records is considered a measure of the success of the match.⁹ This generalization is appropriate for matches where there is every reason to expect that all records on both sides should find a match. That is not the case for this particular match, however. On average, Medicare only pays for about one-half of a beneficiary’s personal medical expenditures (Waldo et al., 1989). Survey reports cover all medical services, not just Medicare covered services. As noted, liberal rules were deliberately used to define what survey reports would be included in the match. In these circumstances, a large number of unmatched survey reports would not be unexpected or surprising.

⁹ See, for example, Cohen’s (1996) discussion of the preliminary results from this match which were presented in Eppig and Edwards (1996).

These unmatched survey reports are predominantly non-Medicare services that should not have matched a Medicare claim.

There are also good reasons for unmatched Medicare claims on the other side. Unmatched Medicare claims, in our matching scheme, can be viewed as a measure of, and a correction for, survey under-reporting. Rather than being viewed as an unsuccessful match, these unmatched Medicare claims are a source of value added to the post-match final file.

WHAT CONSTITUTES AN "EVENT"?

While it is relatively easy to match survey reports and administrative bills at the person level because of the common health insurance claim number in both sources, it is considerably more difficult to match survey reports and billing records at the event level. Often services which are reported as a single event by a sample person are disaggregated into multiple events in the billing records. For example, an outpatient visit may result in multiple Medicare claims. Conversely, multiple visits with the same provider may be reported as separate events by the sample person and be reported on one billing record in the administrative records. What is the best way to construct a one-to-one relationship between survey-reported events and billing data? Should the match be done at the most disaggregate event level possible? Or would a better approach be to match "bundles" of separate services that are conventionally reported together? Using either approach, some basic issues must be addressed.

First, there is wide variation in the resources embodied in the conventional categories used to classify health events, such as inpatient hospital stays, physician visits, outpatient hospital visits, durable

medical equipment, home health visits, etc. These commonly accepted categories of medical "events" differ widely with respect to time covered, resources employed, level of medical skill employed, or therapeutic significance. An inpatient hospital stay, which covers multiple days, involves care from multiple persons, consists of many medical services and supplies, and is very expensive, is considered one event. A simple follow-up visit to a physician is also considered one event. When "events" differ so fundamentally in resource inputs employed and costs covered, meaningful comparisons of events across types of service are very difficult. With regard to matching operations, an unmatched inpatient hospital event is a much more serious matter than an unmatched follow-up visit to a physician.

In addition, as noted in the example of an outpatient visit, a single event reported in the survey may be recorded as multiple events in Medicare billing records (e.g. facility bill, physician services, X-ray, supplies, etc.). Matching these events as they are found in both files means being able to match a single event on one side to a "bundle" of events on the other. These differences in what constitutes an event across service types, and what constitutes an event across the two record sources, make it difficult to find an event definition that is clearly superior, or more appropriate, for the purposes of this match. After considering the alternatives carefully, the decision was made not to always disaggregate to the most fundamental level, but instead to match bundled events in the ways that they naturally occurred in both files. The practical effect of this approach is to concentrate more on getting the charge and payment dollars matched correctly, and less on reconciling differences in how events are recorded in the two sources.

DIFFERENCES IN EVENT CATEGORIES

Another basic difficulty in designing an event-level match between survey-reported events and Medicare billing records is that they are categorized very differently. The MCBS type-of-service categories correspond to the way that an ordinary respondent would classify and group various health services. Medicare billing records, on the other hand, are grouped by the type of provider that furnished the service (Table 2).

There are more than twice as many MCBS categories (9) as Medicare bill categories (4).¹⁰ In some cases this is because Medicare does not cover all medical services, while the survey does. A good example is dental services, which are rarely covered under Medicare. Another category on the survey side that is not shown on the bill side is emergency room services. In the Medicare claims system, emergency room services that are immediately followed by an inpatient stay are included in the inpatient DRG payment. There are no additional separate bills or payments. Emergency room services that do not result in inpatient hospitalization are classified as outpatient hospital services.

EVENT-LEVEL MATCHING

Event-level matching is actually a series of matches. An event from a Medicare claim category must often be matched against more than one MCBS event category, and vice versa. Different algorithms are used in conducting the matches depending on the data elements available. The sequence of matches across categories always proceeds from categories that are

¹⁰ Medicare bill categories also include home health and hospice bills, but these services were matched at the person rather than the event level, and so are excluded from Table 2.

Table 2
Comparison of MCBS Event Categories With Medicare Bill Record Categories

MCBS Event Categories (Classified by Type of Service)	Medicare Bill Categories (Classified by Type of Provider)
Dental (DU)	Inpatient Hospital
Emergency Room (ER)	Skilled Nursing Facility
Inpatient Hospital (IP)	Outpatient Hospital
Outpatient Hospital Services (OP)	Physician/Supplier
Medical Provider Services (MP)	
Other Medical (OM)	
Institutional Utilization (IU)	
Separately Billing Doctors (SD)	
Separately Billing Laboratories (SL)	

NOTE: MCBS is Medicare Current Beneficiary Survey.

SOURCE: MCBS Matching Process for 1992 Cost and Use file.

most likely to match to categories that are less likely. Table 3 shows an overview of the match sequencing.

Matching attempts are done iteratively beginning with strict match criteria and proceeding to less restrictive. For example, reported doctor visits are initially compared with carrier control number, date of service, and total charge. If there is no successful match, the algorithm checks for a match on physician name and date of service or on total charge and date of service. If there is still no successful match, the program looks for a match on physician name and total charge with the date of service relaxed to within a week. The match routines thereby link survey events to Medicare billing records while simultaneously indicating the strength of the link.

As previously noted, the match is designed to allow survey-reported events to be matched to multiple Medicare claims and vice versa. Multiple links are often valid, and the matching process is hierarchical and iterative. For example, a survey-reported doctor visit may be linked to a Medicare bill record for the physician's service and a Medicare bill record for laboratory services for blood drawn during the visit. In some cases, a stronger match occurs later in sequence of matches

Table 3
Overview of Event Category Matches During
Event-Level Matching

MCBS Event Category	Medicare Bill Category
Matches Between Similar Service Types	
IP	Inpatient Hospital
MP, OM, SD, SL	Part B Physician/Supplier
OP	Outpatient Hospital
IU	Skilled Nursing Facility
DU	Part B Physician/Supplier
ER	Outpatient Hospital
Matches Between Less Similar Service Types	
ER	Physician/Supplier
ER	Inpatient Hospital
OP	Inpatient Hospital
IU	Inpatient Hospital
IP	Skilled Nursing Facility
IP	Outpatient Hospital
OP	Part B Physician/Supplier
MP, OM, SD, SL	Outpatient Hospital

NOTES: IP is Inpatient Hospital; MP is Medical Provider Services; OM is Other Medical; SD is Separately Billing Doctors; SL is Separately Billing Laboratories; OP is Outpatient Hospital Services; IU is Institutional Utilization; DU is Dental; and ER is Emergency Room.

SOURCE: MCBS Matching Process for 1992 Cost and Use file.

than an initial weak match. For example, a survey-reported doctor visit may have a weak link to a Medicare physician/supplier record and a strong link to a Medicare outpatient hospital record. MCBS staff used the match strength indicator, and an examination of the potential for bundling and unbundling on both sides, to resolve situations with multiple matches.

This match strategy differs from other approaches, such as that used by the NMES to match medical follow-back provider records to a sample of survey reports (Cohen and Carlson, 1994). In that matching system, statistical probability values are assigned to indicate the strength of a match of survey reports to follow-back provider records. There is an important difference in the objectives of the NMES and MCBS matches, which resulted in different matching strategies. The desired objective of the NMES provider follow-back match is a fully mapped, one-way match of survey reports to the provider follow-back administrative records sample. While a 100-percent match is very difficult to achieve in practice, at

least in theory—or as an ideal objective—there is no reason why each survey-reported event should not find a matching provider record. (There may also be other services from that provider that should have been reported on the survey but were not, but these non-reports do not contravene the point that, at least in theory, 100 percent of survey-reported events should match a provider record.) In these circumstances, non-matched survey reports are regarded negatively as matches that should have occurred, as failures of the matching criteria and processes. In this type of one-way match, a statistical probability value representing potential match strength is a very useful way of characterizing the strength of the link between the survey report and the provider record.

However, the MCBS match is structured differently. As illustrated in Figure 1, there is never any presumption that all survey reports will match a Medicare billing record. The survey collects information on all personal health services, not just Medicare services. This means that unmatched survey reports are to be expected—a health service not covered by Medicare should never match a Medicare billing record. On the other side, it is reasonable to expect, because of memory lapses or lack of full survey participation, that some Medicare billing records will never find a matching survey report. The final file will be composed of three separate elements:

- A file of matched survey-reported events and Medicare billing records in which the best information from each source is combined to make the most complete and accurate record possible.
- A file of unmatched survey-reported events. These are presumed to be non-Medicare covered services.

- A file of unmatched Medicare billing records. These are presumed to be services that should have been reported on the survey, but were not for some reason.

The primary emphasis of the matching processes, in this type of three-way situation, is to be certain that all records are in their correct category. In this matching scheme, any Medicare covered services that should have matched, but did not, will result in duplicate counting when the three segments are combined. There will be an unmatched survey report and an unmatched Medicare billing record that should have been recorded as a single matched event, but instead will be counted as a Medicare non-covered service on the survey side and a survey under-report on the Medicare billing side. In this situation, a single strength of match indicator is less useful than repetitive efforts from different directions to make sure that each record ends up in its proper category. The hierarchical, sequential, and iterative process used for the MCBS match was specifically designed to find all possible matches, and thereby to reduce the risk of double counting in the final file.

For a very large subset of Part B events (around 40 percent) there was a unique carrier claim number available in both the Medicare billing records and the survey-reported event.¹¹ This is the unique claim control number the carrier assigned to the Medicare payment record, and which also appears on the EOMB form sent to the beneficiary. This number, when available, was collected in the survey interview from the EOMB. Because it appeared in both the Medicare billing record and the survey interview reports,

this field guaranteed a correct match for the subset of claims and survey reports on which it appeared.

Cohen (1996) discusses the value of a "truth set," a set of records that are known matches. The records with matched carrier control numbers served that purpose in this match. In addition to the match certainty they provide for a large subset of cases, they also can be used to set and adjust matching criteria for cases where carrier control numbers do not appear. By fine tuning the match criteria using the known matches, these criteria can be set to be sure they do not overmatch (create false positives) or undermatch (create false negatives). One of the more useful insights that came out of the analysis of the known matches was that survey respondents often confused the location of visits, particularly for outpatient hospital and doctor's office visits. Knowing this, the location variable was not relied on as heavily as other variables (such as date of service and doctor's name) in deciding whether there was a potential match. In addition, community physician visits and outpatient hospital visits on both sides were then routinely cross-matched to increase the probability of picking up any misreported potential matches.

After the initial match criteria were established, a person-by-person analysis was conducted. For all persons who had both survey events and Medicare billing records, a determination was made concerning which match criteria resulted in false positives and which match criteria should be relaxed to avoid false negatives. In situations where there are unmatched events on both sides in the same type of service category, more detailed information from the billing records—such as Current Procedural Terminology (CPT) procedure codes—was used to make a

¹¹ Part B refers to the supplementary medical insurance part of Medicare, which covers most medical services other than inpatient hospital and skilled nursing facility care.

judgment whether these items should be a match.¹² This additional service-specific detailed information was often helpful in identifying matches missed in the earlier stages of the matching process.

RESULTS OF THE MATCH

A total of 192,666 Medicare bill events for original sample persons during the time they lived in the community were matched against 179,966 survey reports (Table 4). A match was recorded for 104,349 event records, which is 54 percent of total Medicare bill records and 58 percent of survey-reported events. The percentage of total dollar payments matched was considerably higher. The 88,000 unmatched Medicare bill records represent 46 percent of Medicare events, but only 24 percent of total Medicare payments. The 76,000 unmatched survey events represent 42 percent of all survey events, and 24 percent of survey-reported payments. Looking from either direction, the match was able to account for over three-quarters (76 percent) of reported Medicare payments.¹³

The average payments for unmatched events was considerably lower than for

matched records. Unmatched Medicare events (\$113) were about 60 percent below the average payment for matched events (\$285). This is consistent with past household survey experience that more salient and more expensive medical events are more likely to be remembered and reported at the interview. Unmatched survey reports (\$131) were less than one-half the average payment for matched events (\$285). This is consistent with the fact that Medicare covers the more expensive treatments (such as inpatient hospitalization and outpatient hospital treatment) entered into the match.

The very low average Medicare payments for unmatched survey events (\$23) require some explanation. If the match had worked exactly as hoped, every survey event reporting Medicare dollars should have found a matching Medicare bill record. The unmatched survey events category would consist entirely of non-Medicare services, which by definition should not have any Medicare dollars reported for them (meaning the average Medicare payment in Table 4 should be zero). In fact, about 16,000 of the 75,000 unmatched survey events had a positive Medicare payment amount. The seemingly very low average amount (\$23) results from nearly 16,000 records with reported Medicare payments of \$114 being

Table 4
Summary Results of Matching MCBS-Reported Events to Medicare Bills: 1992

Item	Number of Records	Weighted Record Count (Thousands)	Weighted Total Payments (Thousands)	Average Total Payments	Weighted Medicare Payments (Thousands)	Average Medicare Payments
Survey-Reported Events						
Total	179,966	523,232.7	\$116,202,957	\$222	\$74,935,960	\$143
Matched	104,349	309,803.2	\$88,229,585	\$285	\$70,063,975	\$226
Unmatched	75,617	213,429.5	\$27,973,373	\$131	\$4,871,985	\$23
Medicare Bills						
Total	192,666	556,126.6	\$116,007,497	\$209	\$92,941,204	\$167
Matched	104,349	309,803.2	\$88,229,585	\$285	\$70,063,975	\$226
Unmatched	88,317	246,323.5	\$27,777,913	\$113	\$22,877,228	\$93

SOURCE: Processing counts from development of 1992 MCBS Cost and Use file.

averaged together with nearly 60,000 unmatched survey amounts with zero Medicare dollars. We later discuss how the 16,000 unmatched survey events with Medicare dollars were handled in creating the final file.

EVIDENCE SUPPORTING IMPROVED ACCURACY

One of the primary objectives of the match was to test, and where possible, improve the accuracy of survey reporting. Medicare should have been reported as a payer on 100 percent of the 104,000 survey-reported events that matched a Medicare bill. However, as shown in Table 5, Medicare was only reported as a payer for 81,000, or 78 percent, of survey-reported events. This means that MCBS survey respondents were not aware that Medicare was a payer on one of every five events where Medicare records show that program payments were made. By matching survey reports to Medicare bills, 22 percent of the matched survey-reported events were corrected to make Medicare a payer of record.

Table 5 also shows that, for the 104,000 events where survey reports matched Medicare bills, the Medicare payment amount was only reported on 61 percent of

survey reports. This means that for two of every five events paid by Medicare and matched to a survey event, survey respondents are not able to report the amount that Medicare paid. The match made it possible to fill in the correct Medicare payment for the 39 percent of matched survey reports where no Medicare payment amount was reported.

Another dimension of survey-reporting accuracy that could be checked in the match was how accurately the survey respondent reported the total and Medicare payment amounts, when they reported both these items. As shown in Table 6, both a Medicare payment and total payment were reported on 63,000 of the 104,000 matched records (61 percent). However, there were wide differences between survey-reported amounts and Medicare billing record amounts. Survey respondents consistently overestimated Medicare payments for health services. On average, survey reports were 28 percent higher (\$131) than Medicare payments recorded in administrative billing records (\$102).

Part of the higher survey-reported Medicare dollar amounts could be due to the previously noted differences in the way that services are "bundled" on the survey and in Medicare billing records.

Table 5
Reporting Completeness of Matched MCBS Events, by Type of Service: 1992

Type of Service	Survey-Reported Events							
	Matched Records (Unweighted)		Medicare Reported as Payer		Total Payment Reported		Medicare Payment Reported	
	Number	Number	Percent	Number	Percent	Number	Percent	
Total	104,349	81,056	77.7	81,004	77.6	63,782	61.1	
Dental	52	14	26.9	49	94.2	14	26.9	
Inpatient Hospital	2,844	2,072	72.9	458	16.1	263	09.2	
Institutional Utilization	105	38	36.2	9	08.6	2	01.9	
Medical Provider Services	60,209	45,013	74.8	47,637	79.1	35,164	58.4	
Other Medical	4,534	2,962	65.3	3,478	76.7	2,432	53.6	
Outpatient Hospital Services	16,279	12,808	78.7	9,715	59.7	8,611	52.9	
Separately Billing Doctors	14,674	13,400	91.3	14,106	96.1	12,772	87.0	
Separately Billing Laboratory	5,652	4,749	84.0	5,552	98.2	4,524	80.0	

NOTE: MCBS is Medicare Current Beneficiary Survey.

SOURCE: Processing counts from development of 1992 MCBS Cost and Use file.

Table 6
Comparison of MCBS and Medicare Bill Payments,¹ by Type of Service: 1992

Type of Service	Total Records (Unweighted)	Survey-Reported Events			Medicare Billing Records		
		Average Survey-Reported Medicare Payments	Average Survey-Reported Total Payments	Reported Medicare as Percent of Reported Total	Average Medicare Payments	Average Medicare Approved Total Payments	Medicare as Percent of Total Approved Payments
Total	63,285	\$131	\$239	54.8	\$102	\$144	70.8
Dental	14	\$63	\$137	46.0	\$70	\$100	70.0
Inpatient Hospital	244	\$6,687	\$7,753	86.3	\$5,752	\$6,745	85.3
Institutional Utilization	2	\$6,566	\$7,667	85.6	\$1,012	\$2,450	41.3
Medical Provider Services	35,009	\$60	\$135	44.4	\$55	\$83	66.3
Other Medical	2,423	\$169	\$329	51.4	\$148	\$205	72.2
Outpatient Hospital Services	8,342	\$292	\$370	78.9	\$131	\$206	63.6
Separately Billing Doctors	12,735	\$116	\$313	37.1	\$114	\$163	69.9
Separately Billing Laboratory	4,516	\$79	\$185	42.7	\$61	\$89	68.5

¹Matched events reporting both Medicare and total amounts.

SOURCE: Processing counts from development of 1992 MCBS Cost and Use file

We noted earlier that a single survey-reported visit could appear in Medicare payment records as multiple records; for example, a physician's visit, a lab services fee, and a fee for other medical services and supplies. If all three pieces on the Medicare billing side were not matched to the survey report, this would explain part of the higher survey-reported amount in our subset of matched cases. However, the primary reason that survey respondents overstate the amounts that Medicare pays may be more fundamental. Medicare beneficiaries are probably better informed about the provider's charges than the generally lower cost-based DRG payments for inpatient services and fee-schedule based payments for physician services. They may assume that Medicare pays a higher proportion of the provider than actually occurs.

Survey respondents overestimated total payments even more than they overestimated Medicare payments (Table 6). The average survey response estimate for total payments (\$239) was 66 percent higher than the total payments derived from the Medicare approved payment amount on Medicare billing records (\$144). Total Medicare approved payment amounts include several primary components: Medicare payments, private insurance payments, out-of-pocket payments, and Medicaid payments. A large part of the higher reported total payments from survey respondents may be due to a definitional difference. Total Medicare approved payments are the amounts that are payable under current Medicare law and regulations. These amounts may be from a fee schedule, or limited by law in some way, and are generally lower than provider charges. Therefore, Medicare approved amounts could reasonably be expected to be lower than total payments reported on the survey.

Beneficiaries apparently are not aware of the limits and adjustments that Medicare makes to provider charges in reaching the Medicare approved payment amount. They also may not be aware that supplementary private insurance payments and Medicaid payments are keyed to the Medicare approved payment amount, not the provider's charges. Table 6 shows that beneficiaries consistently overestimate total payments made to medical providers, outpatient hospital services, providers of other medical services, and separately billing physicians and laboratories. Beneficiaries are generally reliable when reporting what they pay out of pocket, but in reporting the remainder of total payments they seem to be assuming that the balance of provider charges (not the generally lower Medicare approved amount) are somehow paid in full by Medicare and the other payers.

In the aggregate, survey-reported Medicare payments overstated the Medicare payment shown in the Medicare bill records by \$5.7 billion; the survey-reported total payment overstated the total payment amount from the Medicare bill records by \$16.4 billion (data not shown). One of the effects of these consistent overestimations is to distort Medicare's share of total payments. Survey reports indicate that Medicare paid 55 percent of total payments for the 63,000 services where both Medicare payment and total payment were reported. Medicare billing records, on the other hand, show Medicare's share of total payments to be considerably higher: 71 percent. Whatever the cause of respondents' propensity to overstate both Medicare and total payments, the match made it possible to correct the systematic payment overestimates that would have resulted if only survey reports had been available.

Evidence of Survey Under-Reporting

In a conventional (100-percent mapped) match in which all survey reports were expected to match Medicare bills, the 88,000 unmatched claims and 76,000 unmatched survey reports would suggest that a large number of potential matches were not identified. However, as noted earlier, the MCBS match structure expects unmatched Medicare bills (which represent events that occurred but were not reported in the survey) and unmatched survey reports (which represent health care services not covered by Medicare and therefore should not match a Medicare claim). Table 7 shows record and dollar counts for all matched and unmatched records by type of service. In general, matched records had higher average Medicare payments (\$226) than unmatched Medicare claims (\$93) and unmatched survey reports (\$23).

One way to assess how many of the 88,000 unmatched Medicare paid bills are under-reports—as opposed to unidentified matches—is to examine the characteristics of the unmatched survey events. A step-down analysis of various categories of unmatched was performed to determine the possible extent of unidentified matches in the 76,000 unmatched survey reports. In general, except for one group of claims, we concluded that a large majority of these events could not be reasonably expected to be undiscovered matches.

Unlikely Matches

- Over 10,000 unmatched survey events were for dental services, which are rarely covered by Medicare.
- Almost 8,000 unmatched survey events had total payments equal to zero. These were very likely parts of bundles of services that were covered in one global

payment on the Medicare claim side, for example, postoperative services which were covered by a global surgery fee. Since finding a match would add no dollars to the matched records group, little energy was expended in trying to rebundle these non-payment records in a match.

- Another 5,000 unmatched survey events were for Medicare HMO enrollees. Virtually all of the Medicare services for these persons are paid through a capitated payment amount and no billing records are submitted to HCFA central files. Consequently, the likelihood is very small that their medical events could ever match a Medicare bill record.
- There were 3,500 unmatched survey events where the sample person was only entitled to Part A or Part B of Medicare, but not both. Therefore a survey-reported service could not reasonably be expected to match a Medicare paid bill record for services for which they were not eligible.
- Another 2,200 unmatched events were provided by the Veterans Administration or in a military installation where no Medicare bill would be expected.
- Over 14,000 unmatched survey events were for other medical services. While Medicare covers durable medical equipment such as wheelchairs and supplies such as oxygen, it does not cover many items in the broad other medical services category such as eyeglasses, hearing aids, heating pads, incontinence supplies, etc. Average Medicare payments for unmatched survey reports of other medical events (\$10) were just a small fraction of average payments for matched events (\$132) and unmatched Medicare claims (\$128) in the same category. This suggests that very few of these records have reported Medicare payments, and most unmatched survey events in this

Table 7
Matched and Unmatched Records by Type of Service and Medicare Payments: 1992

Type of Service	Number of Records	Weighted Record Count (Thousands)	Weighted Total Payments (Thousands)	Average Total Payments	Medicare Weighted Payments (Thousands)	Average Medicare Payments
Matched Medicare Bills and Survey-Reported Events						
Total	104,349	309,803	\$88,229,584	\$285	\$70,063,975	\$226
Dental	52	164	\$12,702	\$78	\$4,986	\$30
Inpatient Hospital	2,844	7,864	\$54,196,242	\$6,892	\$48,222,172	\$6,132
Institutional Utilization	105	270	\$778,516	\$2,882	\$692,764	\$2,565
Medical Provider Services	60,209	184,082	\$13,446,203	\$73	\$8,155,036	\$44
Other Medical	4,534	12,063	\$2,401,746	\$199	\$1,543,779	\$128
Outpatient Hospital Services	16,279	44,809	\$8,918,223	\$199	\$5,825,271	\$130
Separately Billing Doctors	14,674	43,526	\$7,041,381	\$162	\$4,713,201	\$108
Separately Billing Laboratory	5,652	17,026	\$1,434,570	\$84	\$906,767	\$53
Unmatched Medicare Bills						
Total	88,317	246,324	\$27,777,913	\$113	\$22,877,229	\$93
Dental	1	2	\$20	\$9	\$13	\$6
Inpatient Hospital	496	1,242	\$7,549,353	\$6,079	\$6,940,278	\$5,588
Institutional Utilization	94	247	\$761,124	\$3,080	\$655,136	\$2,651
Medical Provider Services	24,511	67,843	\$3,889,867	\$57	\$2,851,676	\$42
Other Medical	5,000	12,712	\$2,075,724	\$163	\$1,676,192	\$132
Outpatient Hospital Services	9,291	25,060	\$3,421,062	\$137	\$2,449,424	\$98
Separately Billing Doctors	20,013	53,749	\$6,842,800	\$127	\$5,421,722	\$101
Separately Billing Laboratory	28,911	85,469	\$3,237,963	\$38	\$2,882,788	\$34
Unmatched Survey-Reported Events						
Total	75,617	213,430	\$27,973,373	\$131	\$4,871,985	\$23
Dental	11,312	36,451	\$4,717,634	\$129	\$24,897	\$1
Inpatient Hospital	359	993	\$5,818,571	\$5,862	\$719,529	\$725
Institutional Utilization	79	197	\$467,252	\$2,373	\$6,451	\$33
Medical Provider Services	34,698	94,531	\$5,688,950	\$60	\$1,474,473	\$16
Other Medical	14,710	41,133	\$4,868,962	\$118	\$428,395	\$10
Outpatient Hospital Services	10,042	27,724	\$5,169,842	\$186	\$1,726,216	\$62
Separately Billing Doctors	3,668	10,147	\$1,025,606	\$101	\$380,653	\$38
Separately Billing Laboratory	749	2,255	\$216,557	\$96	\$111,371	\$49

NOTE: MCBS is Medicare Current Beneficiary Survey.

SOURCE: Processing counts from development of 1992 MCBS Cost and Use file.

category are probably services not covered by Medicare.

In summary, the above items taken together mean that over 40,000 of the 76,000 unmatched survey events either definitely could not, or very likely would not, match a Medicare bill event record. This leaves 36,000 unmatched survey events to be explained.

Likely Undiscovered Matches

There is also a group of unmatched survey events that are very likely to be unidentified matches. Almost 16,000 unmatched survey-reported events report-

ed a dollar amount paid by Medicare. These events are questionable because Medicare billing records represent virtually all payments from Medicare trust funds. Although it is remotely possible that these survey reports are Medicare covered services that somehow are not represented in Medicare billing records, the much more likely possibility is that these are unmatched survey events that should have found a match in Medicare bill records. That is, they are really duplicates for an unmatched Medicare bill record. If they were left in the final file summaries, the total and Medicare dollars reported on these records would duplicate total and Medicare dollars already included in the

unmatched Medicare claims. To avoid duplication in the final file, these records were not included in the file summaries created to represent total and Medicare use and cost figures.

As previously noted, if the match completely succeeded in correctly classifying each unmatched survey report as a non-Medicare service, there would be no Medicare payments shown for unmatched survey events in Table 4. By removing these 16,000 unmatched survey events from the final file, we remove all reported Medicare payment dollars from a match class that, by definition, should not include any Medicare covered services.

Ambiguous Events

This leaves about 20,000 unmatched survey events to be explained. There are many medical services and supplies that Medicare does not cover. For example, physical examinations if the person is well, most alternative medicine services, over-the-counter supplies, etc. We assume that most of these events are non-Medicare services that could not have matched, and thus should be added to the final file.¹⁴

Estimate of Survey Under-Reporting

As discussed, 40,000 of the unmatched survey events were unlikely candidates to match a Medicare billing record; 16,000 events with Medicare payment amounts reported were in fact duplicates that should have matched; and a residual 20,000 records were considered more likely to be non-Medicare services than

¹⁴ The match showed that substantial survey under-reporting exists for Medicare covered services. There is every reason to believe that Medicare non-covered services are similarly under-reported. Any duplicate records (unfound matches) in the last 20,000 of unmatched survey events added to the final file are likely to be considerably fewer than the number that would be required to correct for survey under-reporting of non-Medicare services.

unfound matches, and they were added to the final file. Using these figures, it is possible to compute a range for survey under-reporting of Medicare services uncovered by the match.

Subtracting the 16,000 survey report records that should have matched from the 88,000 unmatched Medicare bills, leaves 72,000 records paid for by Medicare, but without a match from the survey. This suggests that 38 percent (72,000 over 192,000) of medical events paid by Medicare were not reported on the MCBS. The estimated share of dollars under-reported on the survey is smaller because, as previously noted, unmatched Medicare bills had lower payments on average than for matched bills. Using the average total payments for unmatched survey reports and Medicare claims to do the calculations, about 20 percent of total payments were under-reported on the MCBS.

A more conservative estimate would add the 20,000 residual records to the 16,000 to make 36,000 unmatched survey-reported events that could conceivably be unfound matches. Subtracting 36,000 from 88,000 leaves 52,000 Medicare bills that do not have a match in survey events. This implies that 27 percent (52,000 over 192,000) of medical events paid for by Medicare were not reported on the MCBS. In dollar terms, about 15 percent of total payments were under-reported. Whichever estimate is preferred, it is clear that survey under-reporting of medical events is a very serious problem for the Medicare population.

Comparing Match File Versus Survey Results Alone

A final way to evaluate the contributions of the match to the accuracy and completeness of the final file is to compare the post-match results to those that would

have been obtained from the survey alone. Table 8 shows total events, Medicare events, total payments, and Medicare payments from the final matched file compared to the survey file alone.¹⁵ The match greatly increased the number of health service events reported on the survey. Total events were 39 percent higher and Medicare events were 80 percent higher after the match when compared to survey reports alone. Even given the discrepancy in how events are reported between sources, and the wide variation in what constitutes an "event," these represent significant corrections to survey reports.

Total survey-reported payments were lowered 10 percent by the match. Medicare payments, on the other hand, were increased 27 percent by the match. These adjustments were the net effect of survey respondents simultaneously under-reporting Medicare events while overestimating both Medicare and total payments for the events they did report. These large changes in survey-reported health events and payments in the post-match final reconciled file illustrate the value of the match. The post-match file presents a considerably more accurate and complete picture of health services use and costs by Medicare beneficiaries than would have been obtained from survey data alone.

¹⁵ These are post-imputation, final file dollar estimates.

ACKNOWLEDGMENTS

We are indebted to Gerry Adler, Brad Edwards, Dave Gibson, Gary Olin, Kim Skellan, and Dan Waldo for their comments and suggestions on earlier drafts of this report.

REFERENCES

- Adler, G.S.: Medicare Beneficiaries Rate Their Medical Care: New Data From the MCBS. *Health Care Financing Review* 16(4):175-187, Summer 1995.
- Adler, G.S.: A Profile of the Medicare Current Beneficiary Survey. *Health Care Financing Review* 15(4):153-163, Summer 1994.
- Calore, K.A., and Lim, J.: Results of the National Medical Expenditure Survey Household Survey Medicare Record Component Pretest. In F.J. Fowler, ed.: *Health Survey Research Methods: Conference Proceedings*. DHHS Publication Number 89-3447. National Center for Health Services Research, U.S. Department of Health and Human Services, Washington DC. 1989.
- Chulis, G.S., Eppig, F.J., Hogan, M.O., et al.: Health Insurance and the Elderly: Data From MCBS. *Health Care Financing Review* 14(3):163-181, Spring 1993.
- Chulis, G.S., Eppig, F.J., and Poisal, J.A.: Ownership and Average Premiums for Medicare Supplementary Insurance Policies. *Health Care Financing Review* 17(1):255-275, Fall 1995.
- Cohen, S.B.: Discussion of Session on Integrating Survey and Other Data. In R. Warnecke, ed.: *Health Survey Research Methods: Conference Proceedings*. DHHS Publication Number 96-1013. National Center for Health Statistics, U.S. Department of Health and Human Services, Hyattsville, MD. 1996.

Table 8
Comparison of Survey-Reported Data With Post-Match Data From Community Interviews: 1992

Item	Survey-Reported	After Match to Medicare Claims	Change Caused By Claims Match (Percent)
All Events (Millions)	523.2	762.8	203.6 (39)
Medicare Events (Millions)	308.8	556.1	247.3 (80)
Total Payments (Billions)	\$152.3	\$136.7	-\$15.6 (-10)
Medicare Payments (Billions)	\$73.4	\$92.9	\$19.5 (27)

SOURCE: Processing counts from development of 1992 MCBS Cost and Use file.

Cohen, S.B., and Burt, V.L.: Data Collection Frequency Effect in the National Medical Care Expenditure Survey. *Journal of Economic and Social Measurement* 13:125-151, 1985.

Cohen, S.B., and Carlson, B.L.: A Comparison of Household and Medical Provider Reported Expenditures in the 1987 NMES. *Journal of Official Statistics* 10(1):3-29, 1994.

Eppig, F.J., and Edwards, B.: Computer Matching of Medicare Current Beneficiary Survey Data With Medicare Claims. In R. Warnecke, ed.: *Health Survey Research Methods: Conference Proceedings*. DHHS Publication Number 96-1013. National Center for Health Statistics, U.S. Department of Health and Human Services, Hyattsville, MD. 1996.

Gruenberg, L., Kaganova, E., and Hornbrook, M.C.: Improving the AAPCC With Health Status Measures From the MCBS. *Health Care Financing Review* 17(3):59-75, Spring 1996.

Jabine, T.B., and Scheuren, F.: Goals for Statistical Uses of Administrative Records: The Next Ten Years. Presented at the American Statistical Association Meeting. Philadelphia, PA. August 1984.

Lubitz, J., and Riley, J.F.: Trends in Medical Payments in the Last Year of Life. *New England Journal of Medicine* 328(15):1092-1096, 1993.

Mausser, E., and Miller, N.A.: A Profile of Home Health Users in 1992. *Health Care Financing Review* 16(1):17-33, Fall 1994.

Okner, B.A.: Data Matching and Merging: An Overview. *Annals of Economic and Social Measurement* 3(2):347-352, 1974.

Physician Payment Review Commission.: *1996 Annual Report to Congress*. Washington, DC. 1996.

Rodgers, J., and Smith, K.E.: Is There Biased Selection in Medicare HMOs? Unpublished Manuscript. Health Policy Economics Group, Price Waterhouse LLP. Washington, DC. March 14, 1996.

Rosenbach, M.L., Adamache, K.W., and Khandker, R.K.: Variations in Medicare Access and Satisfaction by Health Status: 1991-1993. *Health Care Financing Review* 17(2):29-49, Winter 1995.

Verbrugge, L.M.: Scientific and Professional Allies in Validity Studies. In F.J. Fowler, ed.: *Health Survey Research Methods: Conference Proceedings*. DHHS Publication Number 89-3447. National Center for Health Services Research, U.S. Department of Health and Human Services, Washington DC. 1989.

Waldo, D.R., Sonnefeld, S.T., McKusick, D.R., et al.: Health Expenditures by Age Group, 1977 and 1987. *Health Care Financing Review* 10(4):111-120, Summer 1989.

Winn, D.M., and Walden, D.C.: Validity of Reporting in Surveys. In F.J. Fowler, ed.: *Health Survey Research Methods: Conference Proceedings*. DHHS Publication Number 89-3447. National Center for Health Services Research, U.S. Department of Health and Human Services, Washington DC. 1989.

Reprint Requests: George S. Chulis, Ph.D., Office of the Actuary, Health Care Financing Administration, 7500 Security Boulevard, N3-02-07, Baltimore, Maryland 21244-1850. E-Mail: GChulis@hcfa.gov