United States General Accounting Office

GAO

Report to the Honorable
Ron Wyden, House of Representatives

August 1993

MEDICARE PART B

Reliability of Claims Processing Across Four Carriers

GAO/PEMD-93-27

GAO

United States
General Accounting Office
Washington, D.C. 20548

Program Evaluation and
Methodology Division

B-251298

August 11, 1993

The Honorable Ron Wyden
House of Representatives

Dear Mr. Wyden:

You asked us to assess the methods being used to approve or deny Medicare Part B claims—including whether they are applied correctly and consistently—as well as describe the characteristics of claims denials that are appealed and of those that are reversed. In response to this request, we examined the process that four carriers used to review Medicare claims, focusing on the methods they used to determine the medical necessity of the service and the soundness of those methods. This letter presents our findings on this process. A later report will examine the characteristics of Medicare Part B denials that are appealed and of those that are reversed.

Objectives, Scope, and Methodology

To develop the information in this report, we visited and surveyed four carriers (California Physicians' Services, Northern California; Transamerica Occidental Life Insurance Company, Southern California; Connecticut General Life Insurance Company, North Carolina; Blue Cross and Blue Shield of South Carolina, South Carolina), met with Heath Care Financing Administration (HCFA) officials, and examined documentation provided by HCFA and the carriers.

The information we collected allowed us, first, to identify those methodological components of the claims review process where reliability is needed to ensure the consistent treatment of claims and, second, to determine whether the four carriers we studied had procedures and mechanisms in place that appropriately addressed the issue of reliability. Although our results cannot be generalized to all carriers, they provide important information about the potential of the current Medicare prepayment claims review system to allow inconsistencies in the way Medicare patients are treated.

We conducted our study in June and July of 1993 in accordance with generally accepted government auditing standards.

Summary of Results

The Social Security Act mandates that carriers pay only those Medicare Part B claims that are reasonable and medically necessary. Because HCFA

(the agency in charge of administering this program) does not dictate medical practice, it gives carriers broad latitude in defining the criteria for determining medical necessity. This latitude, in and of itself, provides for some degree of variability in how similar claims are treated across carriers representing different geographic areas. That is, a policy cannot, at the same time, both allow for local variation in what is or is not medically necessary and also produce uniform results.

HCFA policies also encourage carriers to process claims quickly and at low cost. In our study of four carriers, we found that these carriers did indeed process claims rapidly and inexpensively. Level 1 claims examiners, who primarily edited claims for consistency but in some instances made determinations of medical necessity, were expected to process up to 400 claims daily, and level 2 examiners, who exclusively reviewed claims for medical necessity, were typically expected to make determinations on about half that number. The typical educational level attained by claims examiners was high school, with perhaps some college. Given the time constraints under which these claims examiners operated, questions naturally arise concerning their ability to make reliable determinations of medical necessity.

The four carriers we visited medically reviewed an average of 10 percent of the claims they received in fiscal year 1992. However, in 1994, the proportion of medical reviews that HCFA has budgeted carriers to perform will decline to 5 percent. As a consequence, carriers can be expected to reduce their medical reviews correspondingly. Although this HCFA-mandated reduction is aimed at curbing Medicare administrative costs, it is not clear what other effects this change will have. HCFA has not conducted an evaluation of the effect of reducing the number of medical necessity reviews on either the way services are utilized or the carriers' ability to make reliable decisions regarding medical necessity.

We examined three tasks that carriers perform in the course of determining medical necessity to ascertain whether they had procedures in place that would appropriately address relevant reliability issues. With respect to developing medical policy, we found that carriers followed a formalized protocol that allowed for input from the local medical community. With respect to operationalizing medical policy, we found that, although carriers systematically tested their software for errors prior to implementation, they did not have a comparable methodology for testing the interpretability of the operational instructions used by claims examiners to make determinations of medical necessity. (Carriers did,

however, retrospectively assess instructions in order to correct problems.) Finally, we looked at how carriers applied medical policy and the procedures they followed to ensure that examiners reliably followed instructions in making claims determinations. We found that, in addition to audits conducted by HCFA, carriers had internal quality controls that assessed the performance of claims examiners. However, in both instances, the lack of a blind review limited the ability of these methods to detect areas of ambiguity in medical policy.

With regard to the reliability of the system, the human component was clearly the weakest link. Where medical coverage and medical necessity criteria were quantifiable, carriers had often translated such criteria into computer programs. As a method of disposing of claims, computer programs produce consistent results and are also economical. The task of applying less quantifiable criteria, of the type that are involved in making determinations of medical necessity, was assigned to claims examiners. However, because these examiners had to review claims quickly, and in most instances without benefit of a medical background, it may well have been difficult for them to conduct a substantively thorough review.

In summary, the carriers we visited had constructed a system that was able to process a large number of claims very efficiently. However, it is also the case that this system was less well structured for addressing the question of whether medical care is appropriate or not. Moreover, three factors taken together—the time constraints under which determinations for medical necessity were made, the decentralized way in which medical policies were being developed and operationalized, and the weaknesses in some quality control methods being used—raise questions about the system's potential for treating Medicare claimants inconsistently, both within and across carriers.

Background

The Medicare program was authorized by the Congress in 1965 with the passage of title XVIII of the Social Security Act. The program provides health care benefits to persons 65 years of age or older, certain disabled beneficiaries, and most persons with end-stage renal disease. Since its inception, the program has grown considerably: The number of people with coverage increased from 19 million in 1967 to over 35 million in 1992. In fiscal year 1992, the Medicare program paid for health care services for about 96 percent of those eligible. HCFA, within the Department of Health and Human Services, administers the Medicare program and establishes the regulations and policies under which the program operates.

The Medicare program consists of two distinct insurance programs. Part A (Hospital Insurance Benefits for the Aged and Disabled) covers services furnished by hospitals, home health agencies, hospices, and skilled nursing facilities. Part B (Supplementary Medical Insurance for the Aged and Disabled) covers a wide range of medical services and supplies—including physician services, outpatient hospital services, and home health services not covered under Part A, as well as diagnostic laboratory tests, X-rays, and the purchase or rental of durable medical equipment.

The Medicare program cost about $128 billion in fiscal year 1992. Part B payments have recently been growing faster than Part A payments and accounted for about $50 billion of the Medicare expenditures in fiscal year 1992.

Part B coverage requires beneficiaries to pay monthly premiums, meet a $100 deductible, and pay 20 percent of coinsurance. There is no cap on out-of-pocket expenses for beneficiaries under Part B.

In accordance with title XVIII of the Social Security Act, as amended, HCFA contracts with 34 private insurance carriers to process and issue benefit payment on claims submitted under Part B coverage. Carriers are required to process claims in a timely, efficient, effective, and accurate manner. During fiscal year 1992, carriers processed about 550 million Part B claims submitted by nearly 900,000 physicians and suppliers. HCFA policy requires that carriers must approve or deny 95 percent of "clean" claims—that is, claims that do not require additional documentation—within 30 calendar days.[1] In addition, HCFA regulations require that 95 percent of all claims (clean plus all other kinds) must be approved or denied within 60 calendar days.

Carriers are required by regulation to pay only for services that are covered, and to reject or adjust the claim if they determine that the

[1]These figures may be further broken down as follows: (1) 95 percent of electronically submitted claims from participating physicians must be approved or denied within 15 to 17 days, (2) 95 percent of electronically submitted claims from nonparticipating physicians must be approved or denied within 15 to 24 days, (3) 95 percent of all clean paper claims must be approved or denied within 27 to 30 days.

services were "not medically necessary"; in 1992, approximately 8 percent of the dollar amount of denied claims was attributable to this reason.[2]

Section 1862(2)(1)(A) of the Social Security Act provides the general statutory basis for coverage at the same time that it prohibits Medicare payment for services that "are not reasonable and necessary for the diagnosis or treatment of illness or injury or to improve the functioning of a malformed body member." Although the act specifically excludes certain services from coverage—such as cosmetic surgery and routine dental care—it does not provide a comprehensive list of services and equipment that are either covered or excluded from coverage. Rather, the act gives the Secretary of Health and Human Services the discretionary authority to identify medical services that are not medically reasonable and necessary.

HCFA has determined that a number of medical services and items are not covered by the Medicare program. It has been decided, for example, that cellular therapy, acupuncture, intravenous histamine therapy, vitamin B-12 injections to strengthen tendons and ligaments of the foot, and white canes for use by blind persons, are excluded services or items.

How Carriers Review Medicare Claims

The claims review process consists of a combination of computer algorithms and human decisions. Computer algorithms, commonly referred to as prepayment screens, are used as a way of both automatically assessing the validity of claims and channeling certain types of claims to examiners for further review. Review criteria that have straightforward "yes" or "no" answers can often be handled solely by the computer, which makes the final determination. To illustrate, one criterion that is applied to all Medicare Part B claims pertains to beneficiary entitlement: This criterion states that claims involving persons who are not covered by the Medicare program should not be approved, and thus these claims will be denied in all cases. Because this type of determination is unequivocal, it can be made entirely by computer.[3] When applicable, computerized screens of this type provide a quick and reliable way to process claims.

[2]In fiscal year 1992, carriers denied 116 million Part B claims in whole or part (21 percent of all claims processed) for a total of $16 billion (which represented 18 percent of all billed charges). The percentage distribution of dollar amount denied by reason was as follows: duplicate claim (27 percent), service not covered (17 percent), service not medically necessary (8 percent), claimant ineligible (7 percent), missing information (5 percent), rebundled (3 percent), Medicare not primary insurer (3 percent), filing limit exceeded (1 percent), and other (29 percent).

[3]Such computer-generated determinations are termed "auto-adjudications."

Carriers apply six coverage and medical necessity criteria, established by the Social Security Act and HCFA regulations, to determine whether a claim should be paid. (See table 1.)

Table 1: Six Criteria Used in Prepayment Screens

Criterion	Description
Beneficiary entitlement	Used to determine whether the beneficiary is entitled to Medicare coverage—for example, the beneficiary must meet certain age or disability requirements
Claim eligibility	Used to determine whether a claim meets certain basic requirements—for example, the claim must be submitted within a specified time frame and must not be a duplicate claim
Medicare primary coverage	Used to determine whether Medicare—here referred to as the Medicare Secondary Payor (MSP) Program—is the primary insurance carrier; for example, the MSP program is used to determine whether a beneficiary is covered under a spouse's or employer's health insurance plan; in addition, the MSP program reviews trauma codes—for example, massive head and/or kidney damage may indicate that injuries were sustained in an automobile accident, in which case the beneficiary's automobile insurance carrier would be responsible for primary coverage
Reasonableness of charge	Used to determine the appropriate amount of coverage that Medicare will pay claimant—for example, fee schedules for physicians and for durable medical equipment have recently been established by HCFA; these schedules are used to ensure that Medicare pays a reasonable amount relative to resources (for example, physician time and training, malpractice cost, and so on) that were required to perform the service
Medical condition coverage	Used to determine whether the injury or medical condition is covered by Medicare, which assigns a diagnostic code based upon the International Classification of Diseases, 9th Revision
Medical necessity of service	Used to determine whether equipment or services were reasonable and necessary for beneficiary's medical condition—for example, a carrier may determine that a service was unnecessary or that it was unreasonable for a beneficiary to receive treatment from more than one physician

Note: Application of the first five criteria to claims can often be automated by computer. The criterion of medical necessity, on the other hand, more often requires additional human review.

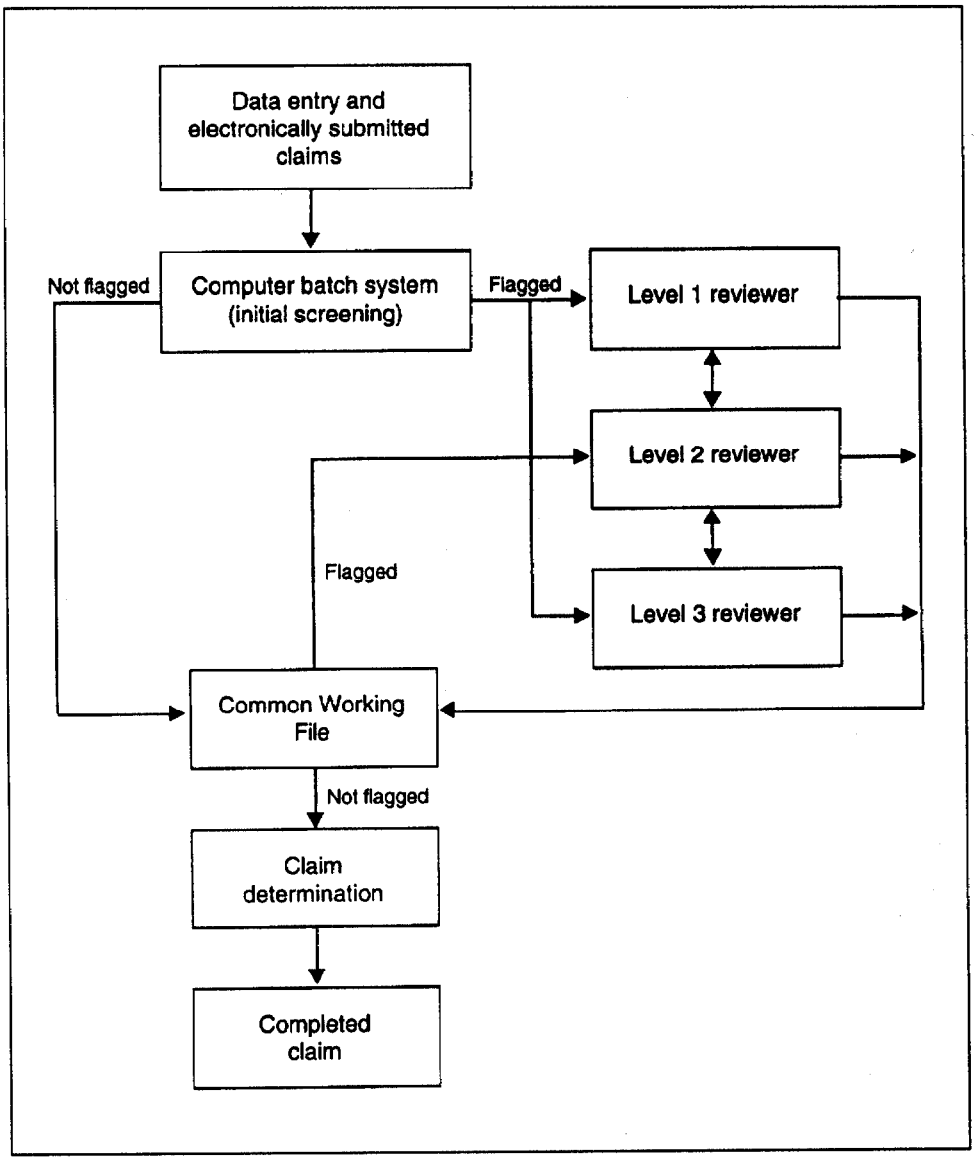
Overview of Claims Processing

Figure 1 depicts, in general terms, how the four carriers we visited structured the process of claims review. After receiving an initial

computer screening, an incoming claim either was routed to the Common Working File (CWF)—where additional automated eligibility checks were made using the beneficiary's past claims history—or, if flagged by a prepayment screen, to one of three levels of personnel who performed a medical review of the claim.[4] Claims flagged for medical review were reviewed by claims examiners using medical criteria developed by the carrier. Claims that could not be determined at one level of medical review might be referred to another level. For instance, if a level 1 examiner could not make a determination on a claim, he or she might pass it on to a more experienced level 2 examiner for determination.

[4]CWF, which is a computer data base maintained for nine regions that contains information on prior claims submissions, is used to determine whether applicable utilization criteria have been met. (For example, a beneficiary is allowed a particular service only once a year.)

Figure 1: Overview of Medicare Part B Claims Processing System



Medical Review

Level 1 examiners primarily performed clerical edits—that is, checked claims for consistency—and in some instances made determinations of medical necessity.[5] The four carriers we surveyed estimated that medical review constituted 5 to 25 percent of the workload of level 1 examiners. In contrast, the task performed by level 2 examiners was solely that of making determinations of medical necessity. Level 2 examiners were usually drawn from the ranks of those at level 1; to attain their position, level 2 examiners usually needed at least 2 to 3 years of prior claims processing experience. It should be noted that while level 1 and 2 examiners may perform what might technically be considered medical review, some tasks that they perform (for example, a comparison of the diagnostic code on the claim with the listed-procedure code) may not require a medical background.

We found a third level of personnel in the carriers we visited consisting of nurses and the medical director.[6] Because nurses command higher salaries than level 1 and level 2 claims examiners, these four carriers employed relatively few medical professionals to assess claims. Two carriers had five and six full-time nurses on staff, respectively; another carrier had one full-time and one half-time nurse on staff, and the last employed one full-time nurse.

The educational requirements for claims examiners were minimal. As shown in table 2, most level 1 and level 2 claims examiners did not have formal medical training; typically, they were high school graduates with perhaps some college experience. The South Carolina carrier was the sole exception: There, level 2 reviews were usually performed by nurses. It should be noted, however, that this carrier did not have nurses performing level 3 reviews—raising the possibility that it defined level 2 review differently from the other three carriers.

[5]Level 1 claims examiners also entered data from paper claims.

[6]HCFA requires that carriers hire a physician to fill the position of medical director. The medical director is responsible for developing medical policy and also may be called upon to examine claims.

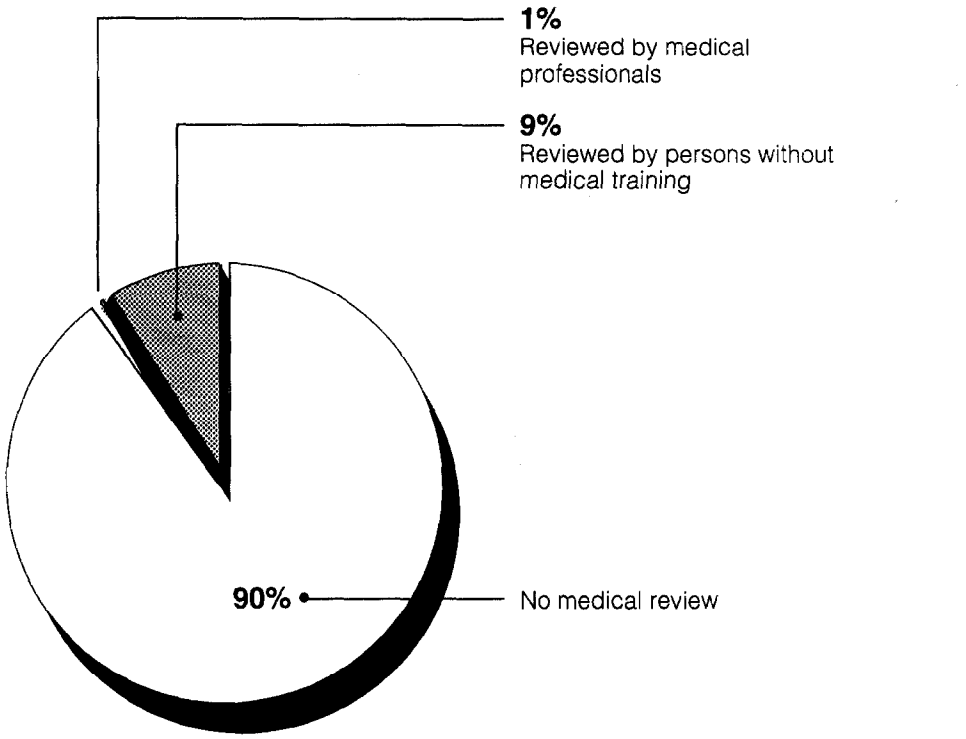
Table 2: Typical Educational Background of Claims Examiners for the Four Carriers

Carrier	Level 1 examiner	Level 2 examiner	Level 3 examiner
Northern California	High school +	High school +	R.N., M.D.
Southern California	High school, college	High school, college + experience	M.D.
North Carolina	High school + 1 year of processing	College or equivalent experience in health-related field	R.N., M.D.
South Carolina	High school + special training	R.N./L.P.N.	M.D.

We found that the four carriers we visited had a tiered screening procedure whereby most claims for which a determination of medical necessity was required were reviewed by claims examiners at levels 1 and 2. As a result of this structure, only a very small number of claims (less than 1 percent) were reviewed by a nurse or medical director. However, to put this figure in context, it is important to recognize that most claims submissions were not flagged for medical review of any type. On average, the carriers we visited medically reviewed 10 percent of the claims they received in fiscal year 1992.[7] As shown in figure 2, in our sample of four carriers, approximately 90 percent of claims were not flagged for medical review by the computer edits, 9 percent were medically reviewed by nonmedical professionals, and fewer than 1 percent were reviewed by nurses or physicians.

[7]By carrier, the percentage of claims medically reviewed in 1992 was Northern California, 9 percent; Southern California, 10.5 percent; North Carolina, 12.9 percent; and South Carolina, 9 percent.

Figure 2: Percent of Claims Submissions Medically Reviewed, for Four Carriers



Note: Percentages are averages for the four carriers surveyed.

As shown in table 3, claims examiners were expected to review a large number of claims each day. The expected rate varied both by level of review and by carrier. Because the complexity of the claims review process increased level by level, level 1 examiners were expected to review more claims per day than level 2 examiners, who in turn were expected to process and review more per day than level 3 examiners. There were also significant differences among carriers; for example, the Southern California and South Carolina carriers expected level 1 reviewers to review about 400 claims daily—a rate equivalent to 50 claims per hour in an 8-hour working day. The Northern California and North Carolina carriers expected level 1 examiners to review about half that number of claims per hour.

Table 3: Expected Number of Claims Reviewed Each Day, by Level, for the Four Carriers

Carrier	Level 1 examiner	Level 2 examiner	Level 3 examiner
Northern California	175	130	18
Southern California	408	200	100
North Carolina	150	35	25
South Carolina	400	100	[a]

[a]Not applicable

National and Local Screens

Beyond identifying services that are not covered by Medicare, HCFA has targeted six medical services (routine foot care, mycotic nails, chiropractic treatments, concurrent care, inpatient rehabilitation medicine visits, and epoetin alpha) for increased scrutiny, and currently requires carriers to develop screens to assess claims for those services. These computerized screens "flag," or direct, some of these service for claims review by claims examiners who, in turn, rely on medical policy guidelines developed by the individual carriers to make determinations. It should be noted that HCFA only identifies services for additional scrutiny and provides certain coverage parameters (for example, number of services allowed per year): It does not provide carriers with the applicable policy guidelines for determining medical necessity—rather, HCFA requires each carrier to individually operationalize this parameter for each of these services.

In addition to the foregoing six nationally mandated prepayment screens, HCFA has granted carriers the authority to develop so-called "local screens" to identify and review other services, which may be of concern in a particular locality, for reasonableness and necessity. HCFA has defined "reasonable and necessary" as meaning

"...whether the service has come to be generally accepted by the professional community as an effective and proven treatment for the condition for which it is being used. If it is, Medicare may make payment.

On the other hand, if the service or treatment is one that is not yet generally accepted, is rarely used, novel or relatively unknown, then authoritative evidence must be obtained to establish it as safe and effective before Medicare may make payment."[8]

In 1988, we reported that the total number of local screens used by carriers ranged from 5 to 177. The four carriers in the present study reported local screen totals ranging from 60 to 91 in number (Northern California, 91 screens; Southern California, 73 screens; North Carolina, 62 screens; and South Carolina, 60 screens).[9]

Reductions in Medical Necessity Reviews

Across the four carriers we visited, the proportion of claims that were reviewed for medical necessity was small (ranging from 9 to 13 percent) in relation to the total number of claims submitted.[10] (See figure 2.) In part, the number of screens—and thus the volume of claims reviewed for medical necessity—is determined by the amount of money that HCFA allots to carriers for the purpose of medical review. Currently, HCFA budgets for up to 9 percent of claims to be manually reviewed for medical necessity but has decided to reduce this level to 5 percent beginning in fiscal year 1994. The carriers told us that, as a result of this change in policy, they intended to turn off some of the screens that assess claims for medical necessity. While this action will certainly reduce administrative costs by eliminating some level 1 and 2 claims examiner positions, the effect that it will have on beneficiary payments and total Medicare expenditures is not known.

HCFA officials stated that the reduction in medical reviews was made to meet cuts in HCFA's appropriation and that this policy was aimed at reducing Medicare administrative costs.[11] They also told us that they have not conducted an evaluation of the potential effect that the reduction of the number of screens will have on utilization of services. Carrier officials we spoke with voiced concern that a reduction in the number of medical reviews might be shortsighted. They noted that any administrative savings

[8]HCFA Intermediary Letter No. 77-4 and 77-5. Medicare & Medicaid Guide (CCH) Para. 28 of 152. Quoted in National Advisory Council on Health Care Technology Assessment, The Medicare Coverage Process (Bethesda, Md.: 1988).

[9]U. S. General Accounting Office, Medicare: Improving Quality of Care Assessment and Assurance, GAO/PEMD-88-10 (Washington, D.C.: May 1988).

[10]It is important to note that a claim may include more than one service, in which case the claims examiner was expected to process each service individually.

[11]HCFA officials have encouraged carriers to conduct more post-payment focused medical reviews and educational programs for physicians to improve billing practices.

that might be achieved by reducing the number of medical reviews could be dwarfed by the costs incurred as a consequence of paying inappropriate claims.

Soundness of Methods of Determining Medical Necessity

Variations Among Carriers

The decision to structure Medicare to allow for variation in local medical practice clearly provided for some degree of variability in how similar claims would be treated by carriers representing different geographic areas. That is, a policy cannot both allow for local variation in what is or is not "medically necessary" and, at the same time, produce uniform results. Thus, one concern is that this method of determining medical necessity may result in too much variability—that is, similar claims may not be receiving sufficiently similar treatment. A second concern is that some screens may be operationally inefficient. Screens with low denial rates are inefficient in that they require disproportionate resources to medically review many claims in order to reject a few. In addition to the associated administrative costs, inefficient screens can impose a burden on providers and beneficiaries who, as part of the process of claims review, may be required to provide additional supportive documentation.

HCFA recognizes that the process by which nationally mandated and local screens are individually developed by carriers—because it reportedly includes input from local medical professionals—may lead naturally to differential health coverage among carriers. This problem is exacerbated in some cases by vague definitions of terms. For instance, carriers "A" and "B" may differ with regard to when they judge the number of claims for a particular service to be "unusually large." HCFA has stated in the Federal Register that "variation is consistent with the legislative intent that the administration of the program take into account both differences in local medical practice and the types of treatment feasible [for] individual patient situations."[12] However, while it may be acceptable for federal policy to lead to some variation in coverage across carriers, local variations can lead to differences in determination of similar claims across carriers.

[12]42 C.F.R. parts 400 and 405; 54 Fed. Reg. 18, January 30, 1989.

Differences Among Carriers in the Number of Claims Denied and Suspended by Prepayment Screens

To examine the extent of variability of claims determinations related to the use of prepayment screens, we analyzed data from HCFA's Medical Review System (MRS) data base for fiscal year 1992. The MRS data base contains carrier-specific information on the number of claims suspended and denied for each of 21 mandated prepayment screens.[13] Our analysis of these data showed that, for the vast majority of prepayment screens, carriers varied markedly in the number of claims they suspended and denied.[14] (See appendix I.)

The nature and magnitude of inter-carrier variation are illustrated by figure 3, which shows the number of claims suspended and denied for one prepayment screen—"mycotic nails"—for each of the four carriers that we visited.[15] It should be noted that, while the workloads of the California carriers were comparable in 1992, those of the North and South Carolina carriers were not.[16]

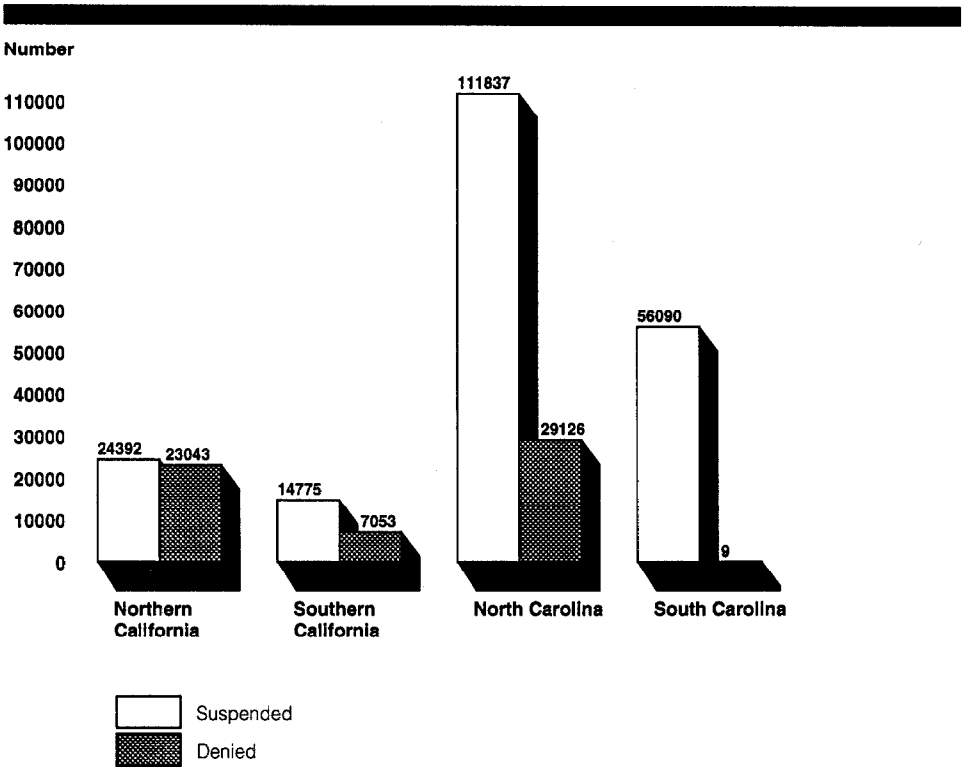
[13]In 1992, HCFA reduced the number of mandated screens—several screens were deactivated and others were classified as optional. A memorandum to regional administrators of Medicare contained the following statement: "Optional screens may be continued, inactivated or revised. It is not necessary to obtain regional office approval to alter or discontinue operational screens." Screens are now mandated only for the following services: routine foot care, mycotic nails, chiropractic services, concurrent care, inpatient rehabilitation medicine visits, and epoetin alpha.

[14]Suspended claims are those claims that are flagged by computer for manual review.

[15]Mycotic nails are a fungal infection of the nails of either hands or feet.

[16]The numbers of claims processed by these carriers in fiscal year 1992 were as follows: Northern California, 23,853,195; Southern California, 24,551,720; North Carolina, 16,606,107; and South Carolina, 5,871,730.

Figure 3: Number of Mycotic Nail Claims Suspended and Denied, for Four Carriers



Note: This graph is meant to illustrate differences in screen efficiency (denied-to-suspended ratio) across carriers. We do not have data on the total number of mycotic nails claims processed by these four carriers. Thus, these figures should not be construed as a comment on the overall denial rate of submitted claims for this service.

Source: HCFA MRS data

Figure 3 shows that the North and South Carolina carriers suspended a larger number of claims for mycotic nail treatments than did the California carriers. A large number of suspensions could be due to one or more of the following factors: (1) there may be regional differences in the prevalence of mycotic nail treatment; (2) carriers may suspect abuse of that service and thus target it for greater scrutiny by suspending more claims; or (3) the screens used by the carriers may not be extensively automated, thus resulting in a greater number of manual reviews. However, in terms of the rate of suspended claims that are eventually denied, a different order emerges. Northern California denies 95 percent of suspended claims, Southern California 48 percent, North Carolina 26 percent, and South

Carolina far fewer than 1 percent. This variability in their denial rates of suspended claims suggests that these four carriers did not employ these mandated prepayment screens in an equally efficient manner.

Carrier Quality Control Methods

Carriers are authorized to determine whether services listed on claims submissions are medically necessary. This entails (1) developing, (2) operationalizing, and (3) applying medical policy. In the following sections, we describe how we used the construct of reliability to identify methodological concerns that carriers should address to ensure the successful performance of each task. In our survey, we asked carriers to describe what procedures they had in place to address these concerns.

Developing Medical Policy

As discussed earlier, HCFA is required by law to delegate to individual carriers the authority for developing the medical policies used to assess claims. Although it might be more economical to have a central body responsible for developing medical policies that then could be used by all carriers, in view of regional differences in medical practices, the current system allows input from the local medical community. Given this local-input objective, then, it is clear that the protocol for developing medical policy should allow for local medical organizations' participation in the process.

To determine whether the procedures used by carriers addressed this concern, we asked carriers to list the steps that they went through to develop a new medical policy. The results of this survey are displayed in table 4.

Table 4: Steps Followed by Four Carriers to Create or Revise Medical Policy Guidelines

Northern California	Southern California	North Carolina	South Carolina
In conjunction with medical director, a new medical policy is proposed. Proposal passes through a series of review levels. Introductions and requests for comments are submitted to carrier advisory committee and all affected medical specialty societies. Organizations have a 45-day comment period. After new policy is finalized, and comments taken into consideration, a 30-day advance notice is published before actual implementation.	Carrier recognizes need for new policy. Consult medical experts and literature, professional associations, data from internal reports, C.P.T. protocol. Evaluate and interpret research data. Formulate policy. Request for comments from carrier advisory committee, professional associations. Organizations have 45-day comment period. After new policy is finalized and comments are taken into consideration, a 30-day notice is given to the medical community prior to implementation. Publication in newsletter bulletin when appropriate. Post-implementation evaluation.	Identify the problem that warrants a new medical policy. Research medical journals for guidelines in developing policy. Formulated policy is presented to associations and/or the carrier advisory committee for comment period of 45 days. After comment period, policy is published in Medicare bulletin. Edit screens are developed to identify services rendered that do not coincide with established policy. These screens are put into production 30 days after policy is published.	Identification of problem, new technology, or HCFA requirement. Identification of codes/specialties involved. Research literature. Draft policy. Send to providers for comment (45 days). Revise as necessary. Publish. Implement.

The approach used to develop a medical policy was generally equivalent among the four carriers we surveyed. The steps followed by carriers can be summarized as follows: (1) identify the need for a new policy; (2) research the area to determine accepted practice (for example, conduct a literature review and talk to experts in the field); (3) formulate the policy; (4) present the policy to the carrier advisory committee comprised of the affected medical specialty societies, which then have 45 days in which to comment; (5) consider and possibly revise the policy based on the committee's comments; and finally, (6) publish the policy 30 days prior to implementation.

The steps followed by carriers to develop new medical policies appear to provide ample opportunity for input from the local medical community. It is unclear, however, whether a decentralized approach to policy development is worth the cost of having carriers independently develop policies for the same services—that is, the amount of regional variation in

medical practice patterns may not be significant enough to justify this approach. HCFA is now in the process of developing a Request for Proposals to establish a clearinghouse of medical policies used by carriers throughout the nation. This clearinghouse, if established, will allow carriers to more easily share information on current policies. It would also allow for a systematic comparison of carrier medical policies, which might shed light on the magnitude of regional variation.

Operationalizing Medical Policy

After a medical policy has been developed, it must be operationalized so that it can be used to determine claims. Operationalizing medical policy involves translating policy statements into computer software and written operational guidelines, which then are used by examiners to make determinations on claims. Operational guidelines contain specific instructions and illustrative examples that aid examiners in interpreting medical policy.[17]

Because it is important that software and operational guidelines accurately reflect medical policy, we asked carriers (1) to describe what procedures they had in place to determine whether the computer software used to screen claims represented a correct translation of medical policy, and (2) to identify the extent to which medical policy rules were clear and interpretable to claims examiners. Carrier responses to these questions are presented in table 5.

[17]Operational guidelines are often referred to as internal medical review guidelines.

Table 5: Actions Taken by Four Carriers to Implement Medical Policy Guidelines

	Northern California	Southern California	North Carolina	South Carolina
How do you determine that the computer software used to screen claims represents a correct translation of medical policy?	Software instructions are reviewed by several different internal divisions (medical review, system support, and the system subcontractor). Once approved, program is tested in the "model office" (test system) to determine whether program will accurately apply policy. This procedure is followed prior to actual implementation.	The software is set up in a test system, and test claims with predetermined results are run through the test system. The results are then analyzed for proper functioning of the software. Adjustments to the software are made if necessary. When the software is functioning correctly, it is placed into production.	Extensive testing of claims against the edits are performed prior to finalizing the screen.	Only policies with clearly set forth criteria (for example, diagnoses, length of stay, and so on) are reduced to computer edits; more complex claims are forwarded to human review.
How do you identify the extent to which medical policy rules are clear and interpretable to claims examiners?	Once a new instruction is established or changed, examiner instructions are prepared. These instructions are reviewed by a network of people, including trainers, quality supervisors and front-line supervisors (and sometimes examiners), to ensure that they are clear, concise, and easily understood. Depending on the nature of the policy, classroom training may be initiated.	Before implementation of a medical policy rule by the claims examiners, a draft is submitted to quality council for clarity and content review. A corrected draft is then submitted to key personnel for clarity and interpretability review. The final implementation rule is then taught to the claims examiners by their section managers.	When new policies are established, draft instructions are issued to all claims supervisors for review and comment. Input from claims processors is obtained, and clarification of instructions is completed if needed. Final instructions are then issued and explained to claims processors.	The medical director is a full-time, licensed physician and is responsible for writing and presenting local medical policy in a manner that is clear and interpretable to both local claims examiners and the medical community. There is no test to determine whether claims examiners find the policies clear and interpretable.

Carriers reported extensively testing new computer software prior to implementation. Software testing involves running a variety of mock claims with predetermined adjudications through the computer. The computer-generated decisions are then compared with claims-examiner decisions to determine whether the software has any errors. If errors are detected, they are corrected before the software is placed into use. This method, properly conducted, appears to be a sound way to ensure the reliability of the computer software used to assess claims.

With regard to the methods used to ensure that written instructions are clear and interpretable to examiners, carriers performed what was essentially a face-validity check. That is, carriers reported having key persons or supervisory staff examine new medical policy instructions for clarity and interpretability prior to implementation. Several carriers stated

that they also solicited input from claims examiners during the preparation of instructions.

Although carriers had a precise methodology to test computer software, using a sample of dummy claims to pinpoint problems, they did not have a comparably rigorous method for identifying ambiguities in the operational instructions used by claims examiners prior to implementation. For instance, the South Carolina carrier reported that "only policy criteria that are clearly set forth (for example, diagnoses, length of stay, and so on) are reduced to computer edits; more complex claims are forwarded to human review." (See table 5.) However, as discussed earlier, since most claims examiners must process claims rapidly without the benefit of a medical background, it would be beneficial to have operational guidelines that are equally straightforward and unambiguous if they are to be used effectively.

In this respect, it appears that pretesting might have provided a more rigorous reliability test of claims examiners' ability to interpret operational instructions. This could be accomplished in a manner analogous to the way in which software is tested. For example, several examiners might be asked to review an identical and representative sample of dummy claims using newly written instructions. If the operational instructions are reliable, the determinations of all examiners should be identical; if they are not, a revision of the instructions might be warranted. Such a test should yield a better estimate of instructional interpretability under operational conditions.

Applying Medical Policy

Proper application of medical policy to actual claims requires that examiners reliably follow medical policy—as explained by the operational instructions developed by a carrier—in making claim determinations. We asked carriers to describe the types of internal quality controls they used to check whether claims examiners were correctly following policy. Carrier responses to this question are displayed in table 6.

Table 6: Actions Taken by Four Carriers to Ensure That Medical Policy Is Reliably Applied

	Action			
	Northern California	**Southern California**	**North Carolina**	**South Carolina**
What types of internal quality controls do you have to check whether claims examiners are correctly following policy?	In-line auditing procedures reveal how medical policy guidelines are being applied. Examiners are reviewed daily, with feedback given weekly. Error trends are charted and refresher training is developed when and where it becomes necessary. HCFA end-of-line quality assurance program ensures quality of processing and medical review through sample auditing of claims.	A quality control section's personnel, on a continual basis, check a set number of randomly selected claims per month (per claims examiner) to see that each examiner is correctly following policy. This is the same method used for the quality assurance system that is required by HCFA—that is, this is a separate but similar system. HCFA end-of-line quality assurance program.	Quality assurance program, using random selection, audits claims for accuracy. Errors that are detected are discussed with the supervisor of the employee making the error. Policy errors are discussed in weekly meetings that include representatives from claims processing, claim exception, medical review, and appeals departments. HCFA end-of-line quality assurance program.	Employ an elaborate quality assurance program designed by HCFA. Each week claims are randomly selected and receive a rigorous quality review by internal staff that are organized separately from the claims examiner staff. HCFA end-of-line quality assurance program.

The carriers' internal, together with HCFA's external, quality assurance programs were cited by all carriers as the principal method for ensuring the reliability of claims processing. This program checks for errors in each of the following six areas: (1) entitlement, (2) coverage, (3) reasonable charge, (4) payment, (5) documentation, and (6) coding or data entry. The quality assurance program uses two methods to estimate claims processing errors. In the first method, each carrier, using a HCFA-developed computer program, selects a random sample of claims processed during the preceding quarter. The carrier then reviews these claims to determine the accuracy of its initial claims processing. Each carrier uses its own set of mandated and local screens to reprocess a claim. Claim determinations from the reprocessed claim are then compared with how the claim was actually decided, to obtain an estimate of error.

The methodology used in the second component of the quality assurance program is similar to the first, except that a 9-percent subsample from the carriers' sample of claims is selected and then reprocessed, this time by HCFA regional staff. Again, decisions are based on the respective medical policies developed by carriers. Estimates of error from the carrier's

sample review are combined with findings from the regional office's subsample review to produce final error rates.[18]

In addition to the above-mentioned methods, carriers reported having internal quality assurance programs. One carrier that we visited had established an in-line auditing procedure that was able to review the performance of claims examiners on a daily basis. Using this system, this carrier was able to provide claims examiners with weekly feedback on the number and types of errors they were making.

To summarize, all carriers had several procedures in place to assess the performance of, and provide feedback to, claims examiners. These assessments were performed on a regular basis by quality control departments within carriers, as well as by HCFA regional staff. If the quality control methods used by carriers are susceptible to criticism, it is with respect to their failure to conduct blind reviews of claims. That is, the internal assessments, as well as those conducted by HCFA regional staff, reprocess a sample of claims to estimate reliability, yet in both instances, those persons reprocessing claims have access to the first examiner's decision. While appropriate for detecting quantifiable errors such as data entry mistakes, these assessments are less well suited for detecting ambiguities in the instructions used by claims examiners.

Agency Comments

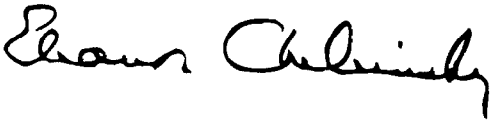
We provided an oral summary of our findings and conclusions to responsible HCFA officials who offered several clarifications of points made in this report, which we have incorporated in the text where appropriate.

Unless you publicly announce its contents earlier, we plan no further distribution of this report until 30 days after the date of this letter. At that time, we will send copies to the Director of the Office of Management and Budget, the Secretary of Health and Human Services, the U.S. Commissioner on Aging, and other interested parties.

[18]HCFA also conducts a Contractors Performance Evaluation Program (CPEP) to evaluate the quality of carriers' claims processing. CPEP does not have any measure of the validity of carriers' medical screens, which affect decisions on medical necessity and appropriateness. See our testimony entitled Medicare: HCFA Monitoring of the Quality of Part B Claims Processing, GAO/T-PEMD-92-14 (Washington, D.C.: September 23, 1992).

If you have any questions or would like additional information, please call me at (202) 512-2900 or Robert L. York, Director of Program Evaluation in Human Services Areas, at (202) 512-5885. Other major contributors to this report are listed in appendix I.

Sincerely yours,



Eleanor Chelimsky
Assistant Comptroller General

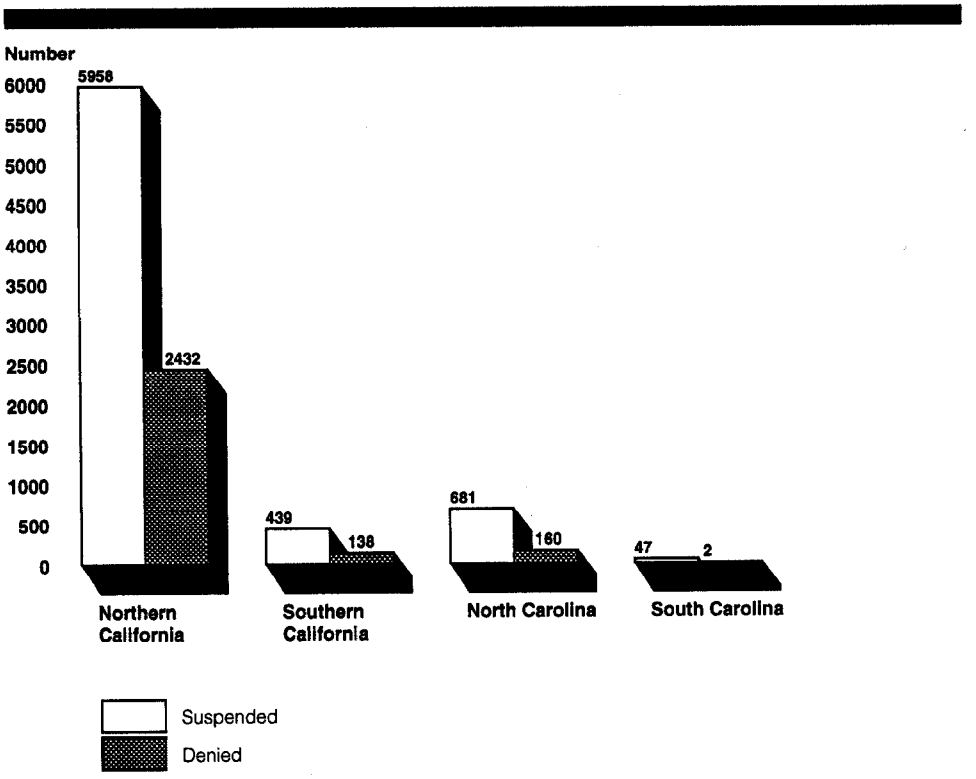
Contents

Abbreviations

CPEP	Contractors Performance Evaluation Program
CWF	Common Working File
HCFA	Health Care Financing Administration
GAO	General Accounting Office
MSP	Medicare Secondary Payor
MRS	Medical Review System

Suspension and Denial Frequencies for Five Nationally Mandated Screens

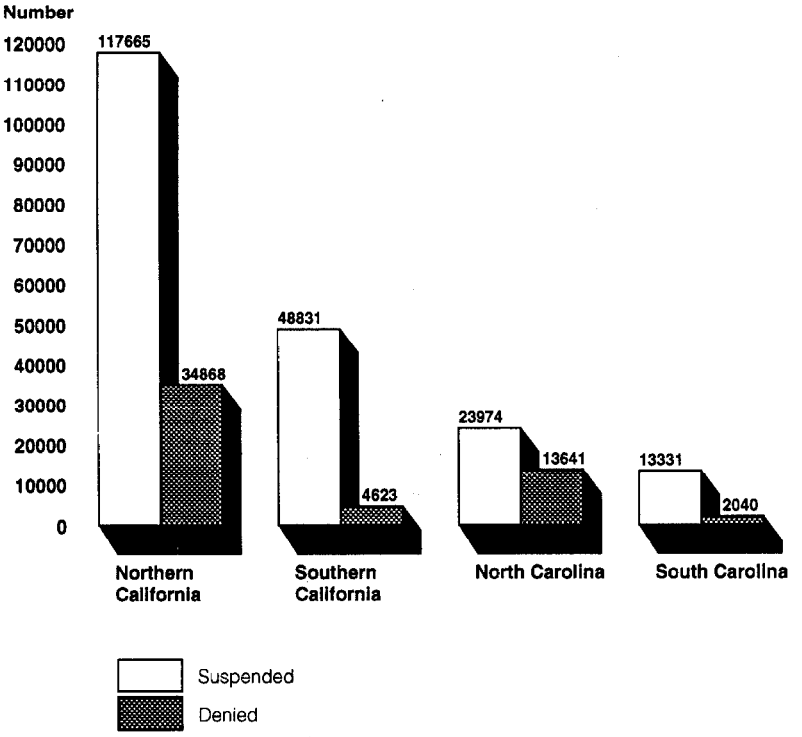
Figure I.1: Number of Routine Foot Care Claims Suspended and Denied, for Four Carriers



Note: This graph is meant to illustrate differences in screen efficiency (suspended-to-denied ratio) across carriers. We do not have data on the total number of routine foot care claims processed by these four carriers. Thus, these figures should not be construed as a comment on the overall denial rate of submitted claims for this service.

Source: HCFA MRS data

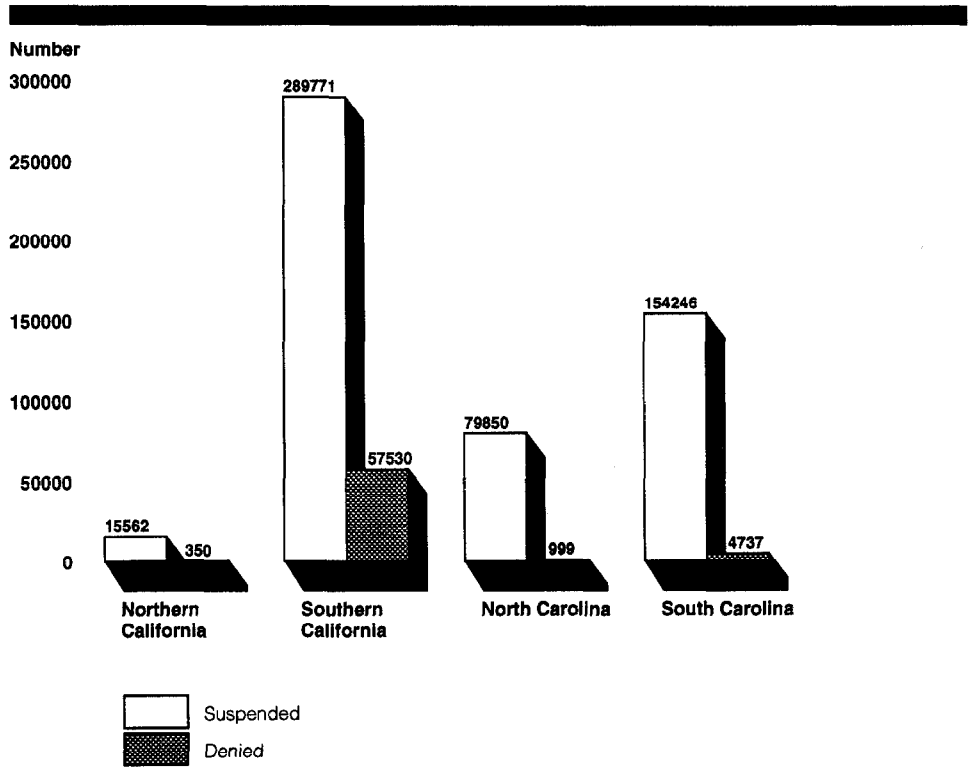
Figure I.2: Number of Chiropractic Treatment Claims Suspended and Denied, for Four Carriers



Note: This graph is meant to illustrate differences in screen efficiency (suspended-to-denied ratio) across carriers. We do not have data on the total number of chiropractic treatment claims processed by these four carriers. Thus, these figures should not be construed as a comment on the overall denial rate of submitted claims for this service.

Source: HCFA MRS data

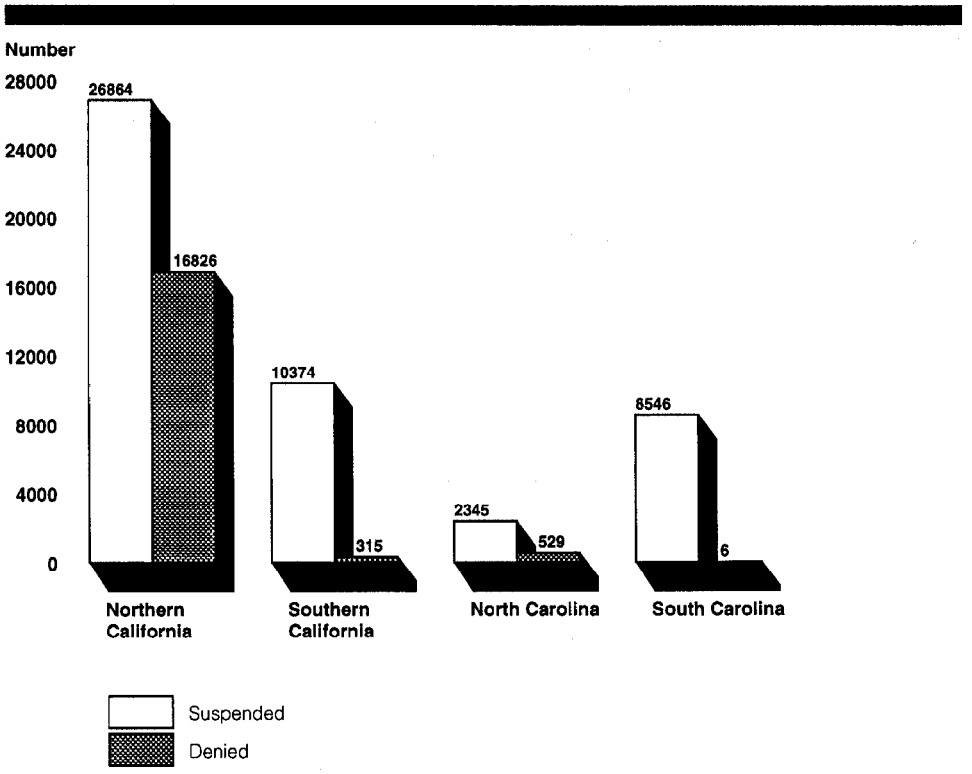
Figure I.3: Number of Concurrent Care Claims Suspended and Denied, for Four Carriers



Note: This graph is meant to illustrate differences in screen efficiency (suspended-to-denied ratio) across carriers. We do not have data on the total number of concurrent care claims processed by these four carriers. Thus, these figures should not be construed as a comment on the overall denial rate of submitted claims for this service.

Source: HCFA MRS data

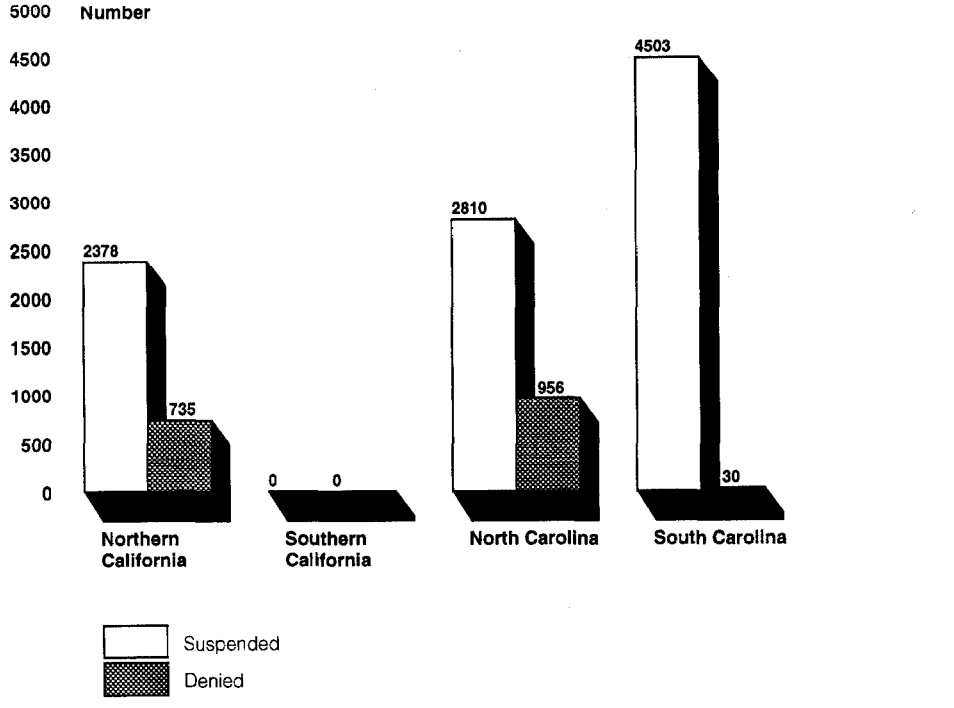
Figure I.4: Number of Inpatient Rehabilitation Visit Claims Suspended and Denied, for Four Carriers



Note: This graph is meant to illustrate differences in screen efficiency (suspended-to-denied ratio) across carriers. We do not have data on the total number of inpatient rehabilitation visit claims processed by these four carriers. Thus, these figures should not be construed as a comment on the overall denial rate of submitted claims for this service.

Source: HCFA MRS data

Figure I.5: Number of Epoetin Alpha Claims Suspended and Denied, for Four Carriers



Note: This graph is meant to illustrate differences in screen efficiency (suspended-to-denied ratio) across carriers. We do not have data on the total number of epoetin alpha claims processed by these four carriers. Thus, these figures should not be construed as a comment on the overall denial rate of submitted claims for this service.

Source: HCFA MRS data

Major Contributors to This Report

Program Evaluation and Methodology Division

Sushil K. Sharma, Assistant Director
Richard M. Lipinski, Project Manager
Brett S. Fallavollita, Advisor
Patrick C. Seeley, Reports Analyst

Ordering Information

The first copy of each GAO report and testimony is free. Additional copies are $2 each. Orders should be sent to the following address, accompanied by a check or money order made out to the Superintendent of Documents, when necessary. Orders for 100 or more copies to be mailed to a single address are discounted 25 percent.

Orders by mail:

U.S. General Accounting Office
P.O. Box 6015
Gaithersburg, MD 20884-6015

or visit:

Room 1000
700 4th St. NW (corner of 4th and G Sts. NW)
U.S. General Accounting Office
Washington, DC

Orders may also be placed by calling (202) 512-6000 or by using fax number (301) 258-4066.

United States
General Accounting Office
Washington, D.C. 20548

Official Business
Penalty for Private Use $300

First-Class Mail
Postage & Fees Paid
GAO
Permit No. G100