
**FY 1985 NATIONAL DRG VALIDATION STUDY:
PEER REVIEW ORGANIZATION CODING ACCURACY**

**RICHARD P. KUSSEROW
INSPECTOR GENERAL**

EXECUTIVE SUMMARY

PURPOSE

This inspection analyzes the accuracy of peer review organization (PRO) coding of diagnosis related groups (DRGs).

BACKGROUND

The "FY 1985 National DRG Validation Study" recoded 7,050 discharges stratified by hospital size. The Office of Inspector General (OIG) located PRO reabstractions for 23.9 percent of the sample. This subsample accurately represented the underlying population by age, sex, diagnosis related groups, and other probable confounders.

FINDINGS

- The PROs correctly coded 78.1 percent of reabstractions, a significantly lower proportion than the 81.8 percent accurately paid by the Fiscal Intermediaries.
- A PRO that changed a DRG had a 90.5 percent chance of being wrong.
- On average, the PROs upcoded their erroneous DRG reabstractions, compounding the financial effect of their lower coding accuracy. These errors caused at least \$172.9 million in overpayments for Fiscal Year 1985.

RECOMMENDATIONS

- The Health Care Financing Administration (HCFA) should determine whether the proportion of PRO reabstraction errors has decreased.
- The HCFA should determine whether the SuperPRO also finds this distribution of changes to PRO coding.
- The HCFA should determine whether internal quality controls of the PROs increase the accuracy of their DRG reabstractions.

In the April 24 comments on a draft of this inspection, the HCFA stated that it has taken a number of actions to improve the accuracy of PRO coding. We agree that these actions respond to the intent of these recommendations. The OIG is now collecting FY 1988 PRO coding data. Comparison of these independently generated data sets will measure the progress achieved by the HCFA in improving PRO coding accuracy.

TABLE OF CONTENTS

| | |
|---|----|
| EXECUTIVE SUMMARY | i |
| TABLE OF CONTENTS | ii |
| INTRODUCTION | 1 |
| BACKGROUND | 1 |
| METHODOLOGY | 4 |
| FINDINGS | 6 |
| PRO REVIEW | 6 |
| PRO CODING ACCURACY | 6 |
| DRG EFFECTS | 8 |
| RECOMMENDATIONS | 10 |
| Appendix 1: PRO review by hospital demography | 11 |
| Appendix 2: PRO review by patient demography | 11 |
| Appendix 3: PRO coding accuracy by hospital demography | 12 |
| Appendix 4: FI coding accuracy by hospital demography | 12 |
| Appendix 5: PRO coding accuracy by patient demographics | 12 |
| Appendix 6: Statistical methodology | 13 |

INTRODUCTION

BACKGROUND

In 1972, Congress established the Professional Standards Review Organizations (PSROs) as part of the Social Security Act. It charged the PSROs with ensuring that all services delivered under Medicare and Medicaid were medically necessary, conformed to appropriate professional standards, and were delivered efficiently and economically. By 1982, however, Congress concluded that the PSRO structure restricted review innovation necessary to limit escalating health care costs and protect against poor quality of care. Therefore, under the Peer Review Improvement Act, Title I, Subtitle C of the Tax Equity and Fiscal Responsibility Act of 1982, Public Law 97-248, Congress created the Utilization and Quality Control Peer Review Organizations (PROs). This legislation required the Secretary of the U.S. Department of Health and Human Services (DHHS) to contract with either "physician-sponsored" or "physician-access" PROs. Each 2-year contract's "scope of work" defined the activities to be performed by the PRO.

Under the first scope of work (1984-1986), the PROs received performance-based objectives to counter potential negative incentives posed by the new prospective payment system (PPS). Under the PPS, hospitals could reduce costs by limiting services and shortening stays, and increase payments by raising the number of admissions or upcoding the diagnosis related groups (DRGs). To control these trends, the PROs had three admission related goals:

- Reduce the number of unnecessary admissions.
- Verify that payment codes conformed to the diagnostic and procedural information in patient records.
- Review cases of transfer or readmission occurring within 7 days of a previous discharge.

The first scope of work gave the PROs four methods for curtailing utilization or quality deficiencies: Education, intensified review, denial of payment, and recommendation of a DHHS Office of Inspector General (OIG) sanction for "substantial violations" in a "substantial number of cases" or "gross and flagrant violation in a single case." Each contract also included five quality of care objectives.

- Reduce unnecessary readmissions.
- Limit preventable complications.
- Reduce unnecessary surgery.

- Lower the risk of mortality.
- Limit avoidable postoperative complications.

To attain these goals and objectives, the PROs received guidelines for extracting data about the physicians and providers in their area. In the first scope of work, these included:

- Five percent random sample of admissions
- Pacemaker cases
- Transfers from PPS hospitals to other hospitals except swing beds
- Readmissions within 7 days of discharge
- DRG reabstraction of DRG 468 and DRG 462 bills
- Statistical outliers

Under the PPS, hospitals receive a pre-established payment for each discharge, based upon the DRG to which the discharge groups. The PPS classified discharges into clinically coherent groups that use similar amounts of hospital resources based on variables such as diagnosis; evaluation and treatment procedures; and patient age, sex, and discharge status. Each of the 473 DRGs had an associated relative weight, which represented the average cost for hospital care provided to patients with diagnoses grouping to that DRG as a proportion of the cost of all patients. The hospital received this payment independent of the actual length of hospitalization or cost of treatment for the individual patient. The hospital retained any surplus from patients consuming less than the expected amount of resources, and absorbed a loss for patients consuming more.

Operationally, the PROs reabstracted PPS bills to check the accuracy of DRG payments to hospitals. The PRO reviewers compared diagnostic and procedural information from the medical record to the narrative diagnoses and procedures attested to the attending physician, and the conversion of narrative information to ICD-9-CM (International Classification of Disease) codes. A registered records administrator (RRA) or an accredited records technician (ART) generally supervised the reabstraction.

During their first scope of work, the PROs claimed to have reabstracted 33.7 percent of PPS bills. Among randomly selected discharges, they found 4.8 percent to be in error. This error rate contrasted with the 9.2 percent reported by the SuperPRO and the 20.8 percent reported by the "FY 1985 National DRG Validation Study: Final Report."

| | |
|--|------------|
| Discharges | 16,547,674 |
| PRO reviews | 5,576,566 |
| [Percent] | [33.7] |
| Source: PPS Monitoring Committee. Report on PPS Monitoring Activities. Feb. 2, 1987. | |

Table 1: PRO reviews under the first scope of work

Sources both inside and outside of the DHHS studied PRO performance under the first scope of work. In the second scope of work (1986-1988), the Health Care Financing Administration (HCFA) adopted many of their findings to strengthen PRO review. These changes included expanding review from readmission within 7 days of a prior discharge to readmission within 15 days of a previous discharge. The PROs also sampled discharges for evidence of premature discharge or transfer. In addition, they reviewed hospitals with unexplained statistical outliers in mortality rates and utilization patterns. The PROs received a standard set of quality screens to apply in all cases selected for review. They also reviewed short stays in the hospital and developed and implemented community outreach programs. All records selected for review were subjected to DRG validation, and the PRO had to have at least one credentialed medical records professional on its staff.

| Review category | Reviews | DRG errors | [%] |
|--------------------|-----------|------------|--------|
| Random sample | 975,280 | 47,535 | [4.8] |
| Intensified review | 319,209 | 18,721 | [5.8] |
| DRG 468 | 160,572 | 18,438 | [11.4] |
| Other | 4,121,505 | 141,900 | [3.4] |
| Total reviews | 5,576,566 | 226,594 | [4.0] |

Table 2: DRG validation reviews for the first PRO scope of work

In developing the third scope of work, the HCFA drew on extensive analysis of PRO activities under the second scope of work. It also incorporated changes mandated in the following public laws passed since the implementation of the second scope of work: Public Law 99-272, the Consolidated Omnibus Budget Reconciliation Act of 1985 (COBRA); Public Law 99-509, the Omnibus Budget Reconciliation Act of 1986 (OBRA 86); and OBRA 87.

The third scope of work remained consonant with the second, but added several significant amendments. To make PRO activities more consistent across contractors, the HCFA developed a basic quality intervention plan, under which the PROs follow a standardized process for categorizing quality problems, notifying providers, and instituting interventions. Also, review of hospital readmissions expanded to patients readmitted within 31 days. Other new requirements include:

- Perform a significant number of on-site reviews in at least 20 percent of rural hospitals.
- Review the reasonableness and medical necessity of invasive procedures.
- Conduct a review of ambulatory surgery procedures.
- Conduct intensified review of those physicians, providers, and DRGs found to exhibit a pattern of substandard care.

- Review post-hospital intervening care for which Medicare payment could be made, that is delivered between two hospital readmissions where the second admission is within 31 days of discharge from the first admission.¹

The HCFA evaluated PRO performance in three ways. To measure PRO activities, the HCFA required monthly and quarterly reports summarizing admission reviews, DRG validations, pre-admission reviews, and other required review data. To evaluate PROs more broadly, the HCFA created the PRO Monitoring Protocol Tracking System (PROMPTS-2). The PROMPTS-2 surveyed PRO performance biannually in such areas as community outreach, management, internal controls, sanctions, and medical review data. Finally, the HCFA contracted with Systemetrics Inc., also known as the SuperPRO, for a biannual review of approximately 400 cases from each of the 54 PRO areas. For each of the cases, SuperPRO repeated the PRO review process using both the generic quality screens and each PRO's own criteria. It compared its results to those of the PRO. The SuperPRO then submitted draft and final reports of its findings to the HCFA. The PROs and HCFA follow-up the SuperPRO's findings in an effort to improve PRO performance.

| # | Cycle | Percent errors |
|---|------------|----------------|
| 1 | 12/84-9/85 | 9.2 |
| 2 | 6/85-2/86 | 9.9 |
| 3 | 11/85-7/86 | 11.5 |
| 4 | 4/86-12/86 | 10.7 |
| 5 | 12/86-9/87 | 10.3 |
| 6 | 6/87-3/88 | 9.3 |

Table 3: SuperPRO reabstraction

This inspection examines one aspect of the PRO review process: DRG validation. The HCFA relies on the PRO validation process to ensure hospital coding accuracy. The hospitals must code their discharges accurately for the PPS to reimburse them fairly and for cost containment. Under PPS, the hospitals forward each bill's ICD-9-CM diagnosis and procedure codes directly to a state Fiscal Intermediary (FI). The FI groups the ICD-9-CM codes to determine the correct DRG, reimburses the hospital, and sends this information to the HCFA. The PROs then review a random sample of these assignments (plus certain specific DRGs) for coding accuracy.

METHODOLOGY

The National DRG Validation Study employed a stratified two-stage sample design based on hospitals and discharges. In the first stage, the OIG used simple random sampling without replacement to select 80 hospitals from each of three bed size strata: Hospitals with less than 100 beds, 100-299 beds, and 300+ beds. The design excluded specialty institutions (e.g., pediatric, rehabilitation, and psychiatric hospitals), facilities in States not using prospective payment at the time (i.e., New York, New Jersey, Massachusetts, and Maryland), and hospitals not contributing data to the calculation of the initial relative weights assigned to diagnosis-related groups. One sample hospital

¹U.S. General Services Administration, National Archives & Records Administration, Office of the Federal Register. *Third PRO scope of work. Federal Register, September 12, 1988; 53 (176): 35234.*

terminated its Medicare eligibility between the study period and actual collection of medical records, leaving a first-stage sample with 239 from a population of 4,913 acute care hospitals.

In the second stage, a systemic random sample selected up to 30 Medicare patients (including persons transferred to other hospitals and those who died) from each of the 239 hospitals for the first half of Fiscal Year 1985. If the hospital discharged fewer than 30 patients during this period, all available discharges were selected. The OIG then

| Bed size | <100 | 1-299 | 300+ | Total |
|-------------------------|--------|-------|-------|-------|
| Hospitals | | | | |
| Population | 2,536 | 1,603 | 774 | 4,913 |
| Sample | 79 | 80 | 80 | 239 |
| Discharges | | | | |
| Population (000,000) | 1.5 | 3.1 | 3.6 | 8.3 |
| Sampled hospitals (000) | 18.2 | 59.5 | 144.7 | 222.4 |
| Sample | 2,276 | 2,388 | 2,386 | 7,050 |
| Sampling fraction [%] | [12.5] | [4.0] | [1.6] | [3.2] |

Table 4: Sampling frame

requested a complete copy of each of the 7,076 medical records selected. With careful follow-up and selective use of subpoenas, the OIG ultimately obtained 7,050 charts (99.6 percent). The sample accurately represented the population of all Medicare beneficiaries discharged during Fiscal Year 1985.

The OIG contracted with the Health Data Institute of Lexington, MA to reabstract the DRGs on a blinded basis. Accredited Record Technicians examined each chart and converted the supportable diagnoses and procedures into ICD-9-CM codes. A Registered Record Administrator supervised the coding team, and contractor physicians provided specialty advice about specific questions. If the reabstracted DRG differed from the DRG paid, a physician evaluated the chart on a blinded basis. A physician panel decided difficult categorizations. Five percent of the entire sample randomly underwent a second, blinded recoding by a different ART to measure the accuracy of the reabstraction methodology. This quality control process revealed no significant discrepancies in the reabstraction process (agreement 0.95, kappa 0.856, Z = 2.12).

The OIG subsequently asked the PROs whether they had reabstracted the chart and if so, the results of the recoding. The PROs reported reabstracting 1,728 of the 7,050 sample discharges, 23.9 percent when weighted by discharges. Smaller hospitals had significantly higher rates of PRO reabstraction (Chi-square 28.5, 2 df, P<0.0001). Appropriately, the "FY 1985 National DRG Validation Study: Final Report" previously noted that smaller institutions have higher rates of billing errors.

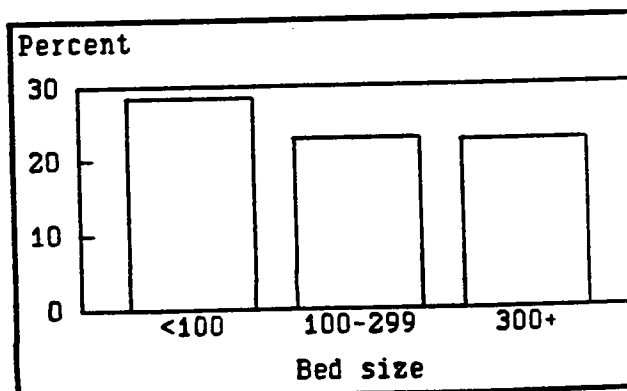


Figure 1: PRO reviews by hospital size

FINDINGS

PRO REVIEW

The PROs disproportionately reabstracted charts from rural (Mantel-Haenszel 56.9, 1 df, $P < 0.0001$) and nonteaching hospitals (Mantel-Haenszel 26.7, 1 df, $P < 0.0001$). As reported by the "FY 1985 National DRG Validation Study: Final Report," hospitals exhibiting these characteristics suffered disproportionate rates of billing errors when controlling for bed size. The PRO selection of discharges appropriately concentrated on institutions with a high yield of reabstraction errors. [Appendix 1].

Patient demographic characteristics for the 1,728 records reabstracted by PROs did not significantly differ from the 5,322 remaining records in average length of stay, mortality rate, or sex distribution. However, the former subsample averaged slightly older than the latter group of patients (t-test 4.5, 1727 df, $P < 0.025$). The PRO sample therefore broadly represented the all Medicare cases. [Appendix 2].

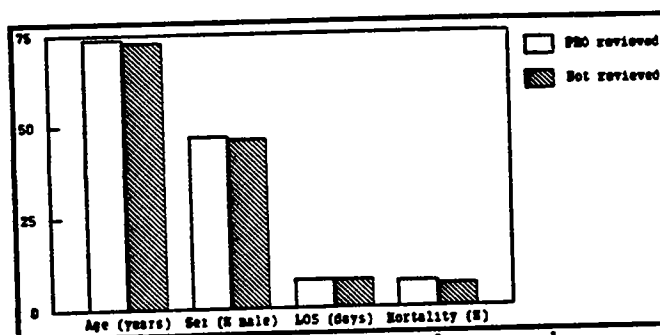


Figure 2: PRO review by patient demography

PRO CODING ACCURACY

When weighted by discharges, the PROs attained 78.1 percent accuracy in their coding reabstractions. Their accuracy improved significantly in larger hospitals (Chi-square 11.2, 2 df, $P = 0.004$). Controlling for hospital size, PRO reabstraction accuracy significantly deteriorated in rural (Mantel-Haenszel 14.7, 1 df, $P < 0.0001$) and for-profit hospitals (Mantel-Haenszel 3.8, 1 df, $P = 0.05$), but not in nonteaching institutions. The PROs increased their efficiency by greater surveillance of rural hospitals, but then decreased their effectiveness with lower coding accuracy in rural hospitals. [Appendix 3].

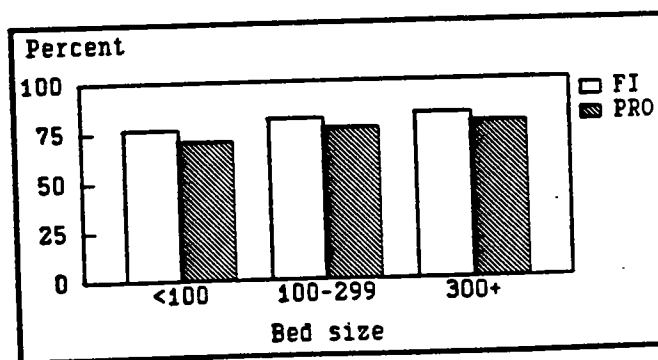


Figure 3: PRO coding accuracy

Interestingly, hospitals code more accurately than the PROs intended to oversee them. The "FY 1985 National DRG Validation Study: Final Report" found that the FIs accurately paid 81.4 percent of bills, weighted by discharges. This proportion

significantly exceeded the PROs' 78.1 percent coding accuracy (Chi-square 25.8, 1 df, $P < 0.0001$). Among the 1,728 discharges reabstracted by the PROs, the FIs accurately paid 81.8 percent, a 1.05 rate ratio. This finding applied across all hospital sizes and demographic characteristics. [Appendix 4]. The subsamples of discharges reabstracted correctly and incorrectly by the PROs did not differ significantly in patient demographics. [Appendix 5].

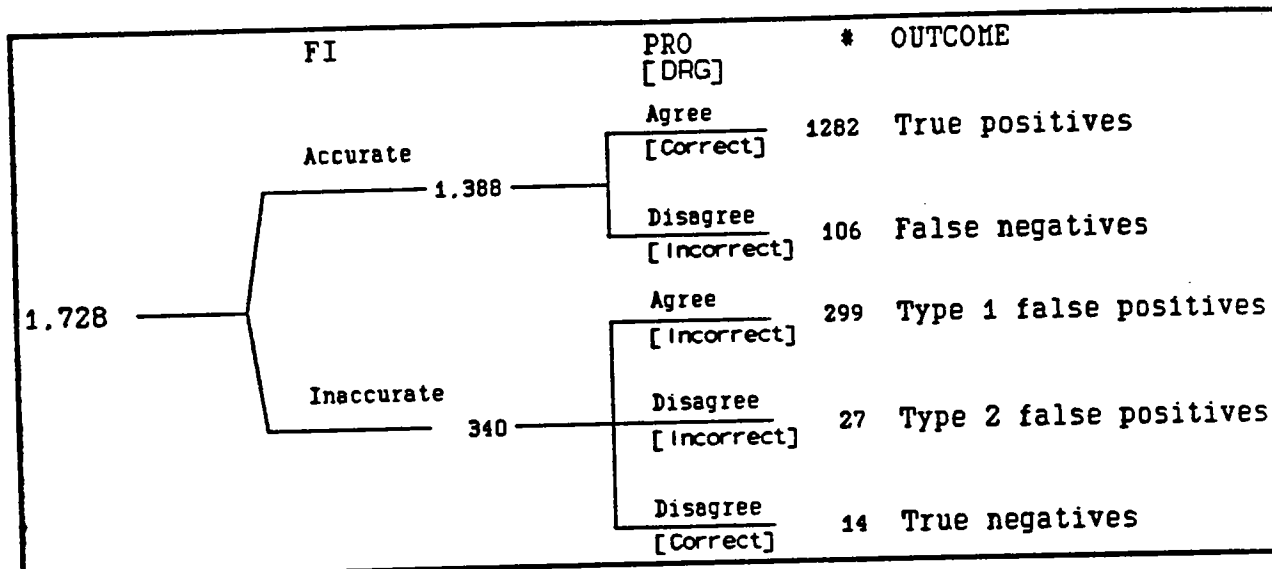


Figure 4: PRO coding results

A PRO's reabstraction of the FI's payment has five possible outcomes. If the FI made the accurate payment, the PRO can either agree (true positive) or disagree (false negative). If the FI paid inaccurately, the PRO can agree (type 1 false positive) or disagree. The latter disagreement may cause the PRO to reabstract the discharge to either the correct DRG (true negative) or to a different, incorrect DRG (type 2 false positive). [Figure 4].

Of this sample's 1,388 bills accurately reimbursed by the FI, the PRO correctly agreed in 1,282 cases; a sensitivity of 92.3 percent. However, in the 340 bills inaccurately paid by the FI, the PRO correctly disagreed and substituted the right DRG in only 14 cases; a 4.1 percent specificity.

| | | Fiscal Intermediary DRG | | Total |
|----------------|-----------|-------------------------|------------|-------|
| | | Accurate | Inaccurate | |
| PRO { DRG { | Correct | 1282 | 14 | 1296 |
| | Incorrect | 106 | 326 | 432 |
| Total | | 1388 | 340 | 1728 |

Sensitivity = 92.3% Specificity = 4.1%

Table 5: PRO coding epidemiology

For this subsample as a whole, FI reimbursement alone would have been correct in 80.3 percent (1388/1,728) of discharges. The PRO review decreased total correct payments to 75.0 percent (1,296/1,728), a 1.07 rate ratio. Under Bayesian analysis, if the PRO makes no DRG

changes, it has 81.1 percent (1,282/1,581) chance of being correct. However, if the PRO elects to change the DRG reimbursed, it has a 9.5 percent (14/147) chance of being correct. The PROs would therefore make a higher proportion of correct decisions by not checking the FI's payment at all. This strategy minimizes both review costs and errors. When the PROs do make DRG changes, they are wrong 10 out of 11 times.

| Outcomes | Number | Relative weight | | | Percent increase (decrease) |
|-----------------------|--------|-----------------|---------|--------|-----------------------------|
| | | FI | Correct | PRO | |
| True positive | 1282 | 1.1157 | 1.1157 | 1.1157 | 0.0 |
| False negative | 106 | 1.1140 | 1.1140 | 1.0707 | (4.0) |
| Type 1 false positive | 299 | 1.0755 | 0.9437 | 1.0755 | 14.0 |
| Type 2 false positive | 27 | 1.2230 | 1.0757 | 1.0498 | (2.5) |
| True negative | 14 | 1.2998 | 1.0247 | 1.0247 | (26.8) |
| Total | 1728 | 1.1118 | 1.0845 | 1.1042 | 1.9 |

Table 6: Relative weights by PRO review

DRG EFFECTS

Not only do the PROs have worse coding accuracy than the FIs, they then compound the financial effect of their mistakes by upcoding. Overall, the relative weights selected by the PROs, whether for correct or incorrect DRGs, exceed by an average of 1.9 percent either the correct relative weights or the relative weights paid by the FIs. Type 1 false positives principally cause the acceleration in overpayment. However, other outcomes also selectively contribute to it.

Of the 106 bills that the FI accurately reimbursed and with which the PRO then incorrectly disagreed, 27.4 percent grouped to only five DRGs. On average for these high frequency DRGs, the PRO incorrectly assigned their bills to higher relative weights than the FI accurately paid. Although this trend did not continue for less frequent DRGs, selective improvement in PRO coding accuracy of false negatives could reverse the net over-reimbursement due to PRO coding errors.

| # | DRG paid | No. | Relative weight | |
|-----|---------------|-----|-----------------|--------|
| | | | Paid | PRO |
| 89 | Pneumonia | 9 | 1.0914 | 0.7834 |
| 127 | Heart failure | 4 | 1.0300 | 1.4150 |
| 138 | Arrhythmia | 4 | 0.9200 | 0.9314 |
| 140 | Angina | 5 | 0.7470 | 0.7305 |
| 182 | Esophagitis | 7 | 0.6121 | 0.9646 |
| | Subtotal | 29 | 0.8842 | 0.9255 |
| | Other | 77 | 1.1868 | 1.1254 |
| | Total | 106 | 1.1140 | 1.0707 |

Table 7: Accurate payments incorrectly changed by the PRO (false negatives)

Similarly, 29.8 percent of the 299 bills inaccurately paid by the FI and with which the PRO incorrectly agreed group to seven DRGs. For each of these DRGs, the PRO also confirmed an average reimbursement higher than the correct DRG. Less frequent DRGs and the subsample as a whole demonstrated the same upcoding. Overall, these

Among 27 inaccurate bills that the PROs changed to different, incorrect DRGs, 48.1 percent group to six DRGs. For these DRGs, the PRO selections also overpay the hospitals to a greater extent than the FIs. Overall, type 2 false positives down code the FIs' relative weight by 2.5 percent. The small number of type 2 false positives and true negatives limits their net effect on PRO upcoding of FI inaccuracies.

| # | DRG | Number | Relative weight | |
|-----|---------------------------------------|--------|-----------------|---------|
| | | | Paid | Correct |
| 14 | Cerebrovascular accident | 10 | 1.3386 | 0.7836 |
| 87 | Respiratory failure | 13 | 1.5368 | 1.0177 |
| 88 | Chronic obstructive pulmonary disease | 11 | 1.0304 | 0.8547 |
| 89 | Pneumonia | 18 | 1.0914 | 0.9827 |
| 127 | Heart failure | 12 | 1.0300 | 0.9548 |
| 132 | Atherosclerosis | 11 | 0.9087 | 0.7727 |
| 182 | Esophagitis | 14 | 0.6121 | 0.8115 |
| | Subtotal | 89 | 1.0704 | 0.8930 |
| | Other | 210 | 1.0776 | 0.9652 |
| | Total | 299 | 1.0755 | 0.9437 |

Table 8: Inaccurate FI payments with which the PRO incorrectly agrees (type 1 false positives)

The PROs' coding errors cost Medicare a minimum of \$172.9 million in Fiscal Year 1985 and potentially larger sums in subsequent years. Prospective Part A payments totaled \$27 billion in Fiscal Year 1985. The PROs reabstracted 33.7 percent of discharges, skewed towards high reimbursement DRGs. The PROs therefore made their 1.9 percent overpayments on at least \$9.1 billion of coding reviews.

| # | DRG | Number | Relative weight | |
|-----|---------------------------------------|--------|-----------------|---------|
| | | | PRO | Correct |
| 14 | Cerebrovascular accident | 2 | 1.5368 | 0.6904 |
| 88 | Chronic obstructive pulmonary disease | 2 | 0.8830 | 0.7088 |
| 89 | Pneumonia | 3 | 1.0229 | 1.1547 |
| 296 | Metabolic | 2 | 0.8003 | 0.7389 |
| 320 | Urinary tract infection | 2 | 0.6121 | 0.6967 |
| 468 | Unrelated procedure | 2 | 0.7016 | 0.9968 |
| | Subtotal | 13 | 0.9336 | 0.8559 |
| | Other | 14 | 1.1577 | 1.2798 |
| | Total | 27 | 1.0498 | 1.0757 |

Table 9: Inaccurate FI payments with which the PRO changes to a different, incorrect DRG (type 2 false positives)

RECOMMENDATIONS

- The HCFA should determine whether the proportion of PRO reabstraction errors has decreased.
- The HCFA should determine whether the SuperPRO also finds this distribution of changes to PRO coding.
- The HCFA should determine whether internal quality controls of the PROs increase the accuracy of their DRG reabstractions.

In the April 24 comments on a draft of this inspection, the HCFA stated that it has taken a number of actions to improve the accuracy of PRO coding. We agree that these actions respond to the intent of these recommendations. The OIG is now collecting FY 1988 PRO coding data. Comparison of these independently generated data sets will measure the progress achieved by the HCFA in improving PRO coding accuracy.

Appendix 1: PRO review by hospital demography

| Number [Percent PRO reabstractions] | <u>Bed size</u> | | | Total | Discharge weighted average |
|---|-----------------|------------|------------|-------|-------------------------------|
| | <100 | 100-299 | 300+ | | |
| Urban | 119 [26.2] | 323 [19.3] | 471 [21.0] | 940 | [21.3] |
| Rural | 528 [29.0] | 222 [31.2] | 65 [45.8] | 788 | [37.2] |
| Teaching | 19 [31.7] | 94 [20.9] | 233 [17.7] | 346 | [21.5] |
| Nonteaching | 628 [28.3] | 451 [23.3] | 303 [28.3] | 1382 | [26.4] |
| Profit | 50 [23.8] | 99 [23.6] | 14 [23.3] | 163 | [23.5] |
| Nonprofit | 597 [28.9] | 446 [22.7] | 522 [22.4] | 1565 | [23.7] |
| Total | 647 [28.4] | 545 [22.8] | 536 [22.5] | 1728 | [23.9] |

Appendix 2: PRO review by patient demography

| | | <u>Bed size</u> | | | <u>Weighted percentage</u> | | |
|------------------|------------|-----------------|---------|------|----------------------------|-----------|----------|
| | | <100 | 100-299 | 300+ | Sample | Discharge | Hospital |
| Age (years) | PRO review | 75.9 | 74.0 | 72.6 | 74.2 | 73.7 | 74.8 |
| | No review | 75.5 | 73.0 | 71.4 | 73.3 | 72.8 | 74.0 |
| Sex (% male) | PRO review | 47.1 | 45.1 | 47.8 | 46.7 | 46.7 | 46.6 |
| | No review | 41.8 | 45.5 | 48.2 | 45.2 | 46.0 | 44.0 |
| LOS (days) | PRO review | 6.0 | 7.2 | 8.7 | 7.3 | 7.6 | 6.8 |
| | No review | 5.8 | 7.5 | 8.3 | 7.2 | 7.5 | 6.7 |
| Mortality (%) | PRO review | 6.3 | 7.3 | 7.3 | 7.0 | 7.1 | 6.8 |
| | No review | 5.4 | 5.9 | 6.9 | 6.1 | 6.2 | 5.8 |

Appendix 3: PRO coding accuracy by hospital demography

| Number Percent correct] | Bed size | | | Total | Weighted percentage | | |
|----------------------------|------------|------------|------------|-------|---------------------|-----------|----------|
| | <100 | 100-299 | 300+ | | Sample | Discharge | Hospital |
| Urban | 87 [73.1] | 245 [75.9] | 398 [79.9] | 730 | [77.6] | [78.1] | [75.1] |
| Rural | 370 [70.1] | 172 [77.5] | 24 [63.2] | 566 | [71.8] | [72.6] | [71.4] |
| Teaching | 11 [57.9] | 72 [76.6] | 188 [80.7] | 271 | [78.3] | [79.1] | [67.6] |
| Nonteaching | 446 [71.0] | 345 [76.5] | 234 [77.2] | 1025 | [74.2] | [75.2] | [73.8] |
| Profit | 38 [76.0] | 68 [68.7] | 6 [42.9] | 112 | [68.7] | [67.0] | [68.4] |
| Nonprofit | 419 [70.2] | 349 [78.3] | 416 [79.7] | 1184 | [75.7] | [77.1] | [74.3] |
| Total | 457 [70.6] | 417 [76.5] | 422 [78.7] | 1296 | [75.0] | [78.1] | [73.8] |

Appendix 4: FI coding accuracy by hospital demography

| Number [Percent accurate] | Bed size | | | Total | Weighted percentage | | |
|------------------------------|------------|------------|------------|-------|---------------------|-----------|----------|
| | <100 | 100-299 | 300+ | | Sample | Discharge | Hospital |
| Urban | 94 [79.0] | 268 [83.0] | 399 [84.7] | 761 | [82.2] | [83.0] | [81.2] |
| Rural | 391 [74.1] | 185 [83.3] | 51 [78.5] | 627 | [78.6] | [79.5] | [77.8] |
| Teaching | 13 [68.4] | 77 [81.9] | 200 [85.8] | 290 | [78.7] | [81.2] | [75.6] |
| Nonteaching | 472 [75.2] | 376 [83.4] | 250 [82.5] | 1098 | [80.3] | [81.5] | [79.0] |
| Profit | 39 [78.0] | 79 [79.8] | 9 [64.3] | 127 | [74.0] | [72.6] | [76.4] |
| Nonprofit | 446 [74.7] | 374 [83.9] | 441 [84.5] | 1261 | [81.0] | [82.4] | [79.2] |
| Total | 485 [75.0] | 453 [83.1] | 450 [84.0] | 1388 | [80.7] | [81.8] | [79.1] |

Appendix 5: PRO coding accuracy by patient demographics

| | Correct | Incorrect |
|---------------|---------|-----------|
| Age (years) | 74.1 | 74.9 |
| Sex (% male) | 46.8 | 46.5 |
| LOS (days) | 7.3 | 6.9 |
| Mortality (%) | 6.3 | 8.3 |

Appendix 6: Statistical methodology

1. $M_p = \sum W_i M_i$
2. $P_p = \sum W_i P_i$
3. $SE_M = (\sum W_i^2 SE_i^2)^{1/2}$
4. $SE_p = (\sum W_i^2 P_i (P_i - 1) / N_i)^{1/2}$

Where:

\sum = summation

M_p = the population mean,

P_p = the proportion of a population with some characteristic,

SE_M = the standard error of the mean,

SE_p = the standard error of the proportion,

W_i = the strata weights,

M_i = the strata means,

P_i = the strata proportions,

SE_i = the strata standard errors.

When calculating statistics for measuring relationships in cross-tabulations BOTEC used the following procedure: "First, we calculated tables separately for each stratum. We then weighted each entry of each table by its stratum weight, summed the entries to get an overall table, and divided each entry in this table by the sum of the stratum weights. The resulting weighted table was used to calculate the strength of sample relationships and the statistical significance of such relationships."

One further complication occurs frequently in BOTEC's statistical analysis of National DRG Validation Study data. "Sometimes the focus of our analysis is upon hospitals, at other times it is upon individual Medicare cases. In the former case the strata weights we use are the inverse strata sampling rates for hospitals, i.e., the ratio of the total number of hospitals in each stratum to the number of hospitals which National DRG Validation Study sampled from that stratum. In the latter case the strata weights used are the inverse sampling rates for Medicare cases, i.e., the ratio of the total estimated number of Medicare cases in each stratum to the number of Medicare cases which National DRG Validation Study sampled from the stratum."