The Need to “Recover”
Recovery Auditing

Karen Matarazzo
Loyola University – Chicago
School of Law
Abstract

The Medicare Trust Fund is at risk of insolvency due to various contributing factors. One important problem is the high number of improper payments made by the Centers for Medicare and Medicaid Services (CMS) each year. CMS implemented the Recovery Audit Contractor (RAC) Program to recover improper payments and return money to the Trust Fund. The RACs were effective at returning money to the Trust Fund (more than $10 billion since the program began\(^1\)) primarily via the audit of acute hospital claims for patient status (specifically, observation or inpatient admission) which carry the highest rates of improper Medicare payments.\(^2\) However, there were some negative consequences. The two most significant disadvantages were: that the provider community was enormously unhappy with RAC activities due to the administrative and financial burden they imposed; and that a backlog of appeals developed at the third level of the Medicare Appeals Process.

CMS instituted several administrative changes, and one crucial regulatory change, which were intended to reduce both provider burden and dissatisfaction and the volume of appeals. These changes altered the types and numbers of claims that the RACs can audit. Unfortunately, these modifications also brought such significant changes to the RAC program that only a small fraction of the monies previously recovered by the RACs are being returned to the Trust Fund.

Regulatory change is needed so that the CMS RACs can, again, effectively return monies to the Trust Fund. CMS should: clarify the guidance pertaining to the hospital admission decision by defining and codifying observation as an admission status; and reestablish RAC auditing of acute hospital patient status claims to increase the amounts being returned to the Trust Fund via recovery of improper payments. To improve the appeals process, CMS should: codify the RAC Discussion Period to make it a mandatory part of the RAC review process, thereby shifting the burden to RACs and reducing the number of appeals entering the Medicare Appeals System; and make Local Coverage Decisions mandatory authority to improve consistency of Administrative Law Judges (ALJ) decisions on appeal.


\(^2\) Ibid.
I. Introduction

The Medicare Trustees estimate that the Medicare Hospital Insurance Trust Fund will be bankrupt by 2030. The number of Medicare beneficiaries is projected to climb by 36% by the mid-2020s. Medicare spending continues to grow at alarming rates, with projections to keep increasing annually for the next decade. Healthcare expenditures due to improper payments, specifically, are a substantial drain to the Trust Fund. Improper payments are any payments that should not have been made, or that were made in an incorrect amount, under statutory, contractual, administrative, or other legally applicable requirements. Errors in payment quickly reach sums in the billions due to the size and scope of the Medicare program.

From 1996 through 2002, the Office of the Inspector General (OIG) provided a general estimation of the Medicare Fee-For-Service (FFS) improper payment rate. With the enactment of the Improper Payments Information Act of 2002, the Centers for Medicare and Medicaid Services (CMS) assumed responsibility for determining the FFS improper payment rate from the OIG. The first-ever *Improper Medicare Fee-For-Service (FFS) Payments Report* published in 2003 noted a national improper payment rate of 9.8%, representing over $19 billion in erroneously paid claims. CMS went about initiating corrective actions to achieve one of its performance goals to lower the rate of improper payments to 5% or less. Also in 2003, the Medicare Modernization Act (MMA) was passed. The success of recovery auditing in the commercial insurance industry was recognized and the MMA included verbiage that directed the creation of a pilot program to determine if Recovery Audit Contractors (RACs) could efficiently and effectively perform the same function for CMS.

---


4 Ibid.

5 Ibid.


7 Ibid.


12 Id at p30.


CMS’s corrective action initiatives also included educational programs, improved communication, and contractor-based medical review strategies.\(^{15}\) As a result, the FFS improper payment rate improved by 2005 to 5.2%.\(^{16}\) However, this still represented a significant fiscal impact: more than $12 billion in improper payments.\(^{17}\) The three-year RAC demonstration program commenced in 2005.\(^{18}\) RACs began to review Medicare FFS claims for improper payments in three states, recouping the government’s money for any findings of over-payments, and returning money to providers for findings of under-payments.\(^{19}\) Within two years, hundreds of millions of dollars were returned to the Trust Fund and the program was expanded to a total of six states.\(^{20}\) The rate of FFS improper payments continued to improve to 3.7% by 2008.\(^{21}\) Based on the success of the demonstration, the permanent RAC program was created in 2009 and this included a mandate to expand the program to all states by 2010.\(^{22}\) Although improvements\(^{23}\) were instituted in the national program based on issues identified during the demonstration program, the *raison d’être* for the RACs remained the same, to safeguard the Trust Fund by identifying and recovering improper payments.\(^{24}\)

During the first RAC contract, the focus of the audit was on patient status reviews for acute hospital claims, which is an area that causes some of the highest rates of improper Medicare payments.\(^{25}\) In fact, between 2010 and 2017, an average of approximately 75% of improper payments collected came from inpatient hospital claims.\(^{26}\) The RACs audited claims for medical necessity; the claims were for services that would have been appropriate, clinically, if they had been provided the less intense setting of outpatient observation as opposed to inpatient


\(^{17}\) Ibid.


\(^{23}\) For example: Medical directors and coding experts were optional for the demonstration, but mandated for the permanent national program.


admission. The audit of observation claims was a significant source of revenue returned to the Medicare Trust Fund: $1 billion during the demonstration project alone.

Although the RACs brought significant financial benefits to the Medicare program, there were detrimental effects as well. One chief problem was dissatisfaction in the provider community. Providers were unhappy with the administrative and financial burden the RAC program brought, from initial review through appeal. Providers also contended that the RACs were inaccurate in their recovery determinations. Another significant detriment was the number of appealed claims.

There are five levels to the Medicare Appeals process, four of which are administrative. The first two levels are performed by Medicare contractors: the Medicare Administrative Contractors (MAC), and the Qualified Independent Contractors (QIC). The third level of appeal consists of a hearing before an Administrative Law Judge (ALJ); at the fourth level, appealed claims are reviewed by the Medical Appeals Council; and the final level is judicial review before Federal District Court. Although the number of appeals increased across all levels, appeals to the third level caused the most problems. The volume became so large that the Office of Medicare Hearings and Appeals (OMHA) could not keep up, resulting in a large backlog of pending appeals, and ALJ decisions rendered long after the statutory time frames.

Due to the negative consequences of the RAC reviews, CMS effectively “hit the pause button” on the RAC program. While several changes were made, the most impactful was that RACs were prohibited from reviewing hospital patient status claims in August of 2013 with the promulgation of the so-called Two Midnight Rule (2MN). The rule changed the benchmark that physicians use as a guide for making an admission decision from 24 hours to 48 hours – or two midnights. The objective of the 2MN Rule was to reduce confusion regarding the inpatient admission decision, and to align the conflicting interests of providers and contractors. The changes enacted were intended to mitigate the disadvantages of the RAC program, however, the effects have been questionable.

---

30 Id at p231.
32 Ibid.
33 Ibid. Original
37 42 CFR 412.3.
There are many who argue the 2MN Rule did little to either improve the clarity of the regulations guiding the admission decision, or to improve the accuracy of those decisions. Additionally, the rate of improper payments remains unacceptably high. The Office of Management and Budget noted that the FFS program had the highest rate of improper payments across government agencies from 2010 to 2016 (see table 1).

Table 1: Medicare FFS Improper Payment Data
Sources: HHS Agency Annual Reports and Medicare FFS Improper Payments Reports

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>National Improper Payment Rate</td>
<td>10.5%</td>
<td>8.6%</td>
<td>8.5%</td>
<td>10.1%</td>
<td>12.7%</td>
<td>12.1%</td>
<td>11.0%</td>
<td>9.5%</td>
</tr>
<tr>
<td>Dollars Spent in Improper Payments (in billions)</td>
<td>$34.3</td>
<td>$28.8</td>
<td>$27.4</td>
<td>$36.0</td>
<td>$45.8</td>
<td>$43.3</td>
<td>$41.0</td>
<td>$36.2</td>
</tr>
</tbody>
</table>

Because any incorrectly dispersed payment will negatively impact the long-term viability of the Trust Fund, improper payments remain a priority concern for the OIG and for CMS. However, the changes imposed on the RACs – most notably the loss of patient status reviews – has debilitated RAC audit and return of monies to the Trust Fund, as illustrated in Figure 1.

---

CMS processes more than a billion Part A and B claims per year. The authorities that dictate how to bill claims for inpatient admissions or outpatient observation services are often found to be overlapping, vague and inconsistently applied by all stakeholders, including the ALJs. Even with the highest levels of integrity and most ambitious efforts, mistakes will be made and improper payments will occur. Regulatory changes are needed to rectify this situation. CMS should: clarify the guidance pertaining to the hospital admission decision by defining and codifying observation as an admission status; and reestablish RAC auditing of short stay hospital claims to increase the amounts being returned to the Trust Fund via recovery of improper payments. To improve the appeals process, CMS should: codify the RAC Discussion Period to make it a mandatory part of the Medicare Appeals Process; and make Local Coverage Decisions binding on Administrative Law Judges (ALJ).

Part II of this paper will further examine the RAC program, its methods, and its pros and cons. Part III will take a deeper dive into the government’s attempts at correcting some of the issues resulting from RAC reviews and why these efforts are inadequate. Part IV explores recommendations for further regulatory and policy changes to address the continuing deficiencies and pain points.

---

46 Id at p3.
II. The RAC Program

A. Methods

CMS has the country divided into regions, and contracts were awarded to contractors by region. RACs are tasked with performing post-payment reviews of provider claims to identify and recover improper payments. RACs may potentially review a variety of claim types, but improper payments can be identified based on four overarching billing inaccuracies: incorrect payments; non-covered services (including services that are not reasonable and necessary); incorrectly coded services (including DRG miscoding); and duplicate services. RACs do not select or perform claim reviews at random and are, in fact, prohibited from doing so. Rather, RACs use proprietary data analysis techniques to find claims that are determined to have a high likelihood to contain payment errors, and then perform targeted reviews of those claims. Any “new issue” that a RAC develops based on a discovered billing inaccuracy must be approved by CMS. Once CMS approves it, the RAC must post a description of the new issue on its website for fifteen days. Then the RAC is permitted to send requests for medical records (known as additional documentation request or “ADR” letters) to providers. When RACs do send ADR requests, they are limited to a three-year look back period from the date the claim was paid.

RACs are instructed to use both automated and complex reviews to find overpayments. Automated reviews are based on clear and unambiguous CMS coverage and payment policy. Claims are discovered through data-mining and no medical record submission or review if required. Complex reviews, on the other hand, require evaluation of medical record documentation by a human (i.e. the appropriate professional subject matter expert: nurses, coders, therapists).

As far as governance, CMS gives the RACs a Statement of Work which contains compulsory directives for all RAC tasks and responsibilities. Additionally, RACs are required

---

47 For the purposes of this paper, discussion will be limited to RACs for Medicare Parts A and B.
49 CMS. Statement of Work (SOW) for the Part A/B Medicare Fee-for-Service Recovery Audit Program – Regions 1-4. at p1.
51 CMS. Statement of Work (SOW) for the Part A/B Medicare Fee-for-Service Recovery Audit Program – Regions 1-4. at p16.
54 Ibid.
55 Ibid.
56 Ibid.
57 CMS. Statement of Work (SOW) for the Part A/B Medicare Fee-for-Service Recovery Audit Program – Regions 1-4. at p17.
58 Id at p22.
59 Ibid.
60 Ibid.
to comply with several forms of regulatory guidance which include: CMS manual provisions; National Coverage Decisions (NCDs); national coverage and coding articles; and Local Coverage Decisions (LCDs), and local coverage and coding articles in their respective jurisdictions. In situations where there is no national or local Medicare policy, the RACs review claims based on accepted medical standards and practice at the time the claim was submitted.

Once reviews are completed and providers are notified of improper payments (under or over-payments), providers have 30 days to initiate a “discussion period.” Providers can use this time to submit additional documentation to support the billing of the claim. RACs must respond with a written, detailed rationale of the discussion determination. This discussion period is separate from and mutually exclusive of the Medical Appeals Process. If a claim is forwarded to the first level of appeal, the RAC must notify the provider immediately that the request for discussion is invalid.

B. Benefits

CMS deemed the RAC program to be cost-effective, as the expense of the demonstration program was considerably less than the Medicare revenue returned. It is estimated that the RAC program costs about twenty cents on the dollar to operate. The ability of the RACs to return money to the Trust Fund is evident: since their genesis in 2010, RACs have recovered more than $10 billion in improper payments and the Chairman and Ranking Member on the U.S. Senate Special Committee on Aging credited the RACs with extending the life of the Medicare program. According to HHS, there has also been a sentinel effect: because of the potential for a RAC audit, providers are more careful about billing accuracy.

C. Disadvantages

Since RACs are private, profit-driven companies, and are paid on a contingency basis, many fear that auditors might be biased in their recovery decisions. There are arguments that

---

64 CMS. Statement of Work (SOW) for the Part A/B Medicare Fee-for-Service Recovery Audit Program – Regions 1-4, at p30.
65 Ibid.
67 Ibid.
RAC scrutiny promoted the (often inappropriate) overuse of observation (instead of inpatient admission) by hospitals wishing to avoid denial for erroneous hospital admissions.\(^3\)

Another unintended consequence of the RAC program is the increase in costs for beneficiaries. Part B services often carry higher copays and co-insurance amount for beneficiaries than Part A services.\(^4\) So, in a case where a Part A inpatient hospital stay is denied as medically unnecessary and the provider rebills under Part B, the out-of-pocket costs may be higher for the beneficiary.\(^5\) A compounding factor is that, in a patient’s eyes, a hospital stay looks the same regardless of how it is billed.\(^6\)

Additionally, in order for Medicare to cover a stay in a skilled nursing facility (SNF), the patient must have a three-day hospital stay preceding the SNF admission.\(^7\) As an outpatient service paid under Part B, observation does not qualify towards admission.\(^8\) Therefore, when a claim for inpatient admission is denied due to lack of medical necessity, even if the hospital rebills the claim under Part B, it does not count towards a subsequent SNF stay.\(^9\) Ultimately, the beneficiary is held liable for the SNF stay which results is high out-of-pocket costs.\(^10\)

Appeals, generally, is another problematic area related to the RAC program. For many hospitals, the appeals process itself is burdensome financially, administratively, and in terms of time spent.\(^11\) The provider community argues that the RACs often recover improperly, asserting that claims are overturned on appeal 72\% of the time.\(^12\) However, the statistics are inconsistent, and the government reports differ from third-party reports. The OIG notes that only 6\% of overpayment claims were appealed, and of those claims, only 44\% of those denials were overturned on appeal.\(^13\) The Government Accountability Office (GAO) noted that the rates of overturn for RAC overpayment claims ranged between 68.0\% and 52.5\% between 2010 and 2014.\(^14\) In its 2016 State of the RAC Program report, the Council for Medicare Integrity characterized rates of overturned RAC overpayment determinations as “low.”\(^15\)

---


\(^{6}\) Id at p11.

\(^{7}\) CMS Internet-Only Manual (IOM) 100-02, Medicare Benefit Policy Manual (MBPM), Ch.8 §20.1. (Rev. 242, 03-16-18).


\(^{9}\) Ibid.


\(^{12}\) See, for example, Jessica L Gustafson and Abby Pendleton. *Billing for and Appealing Denials of Inpatient Hospital Services: Where Have We Been? Where Are We Now? What Does the Future Hold?* (2013) 26 No. 2 Health Law 1 at p8.


\(^{14}\) Government Accountability Office (GAO). Report to Congressional Requesters: *Medicare Fee-for-Service Opportunities Remain to Improve Appeals Process.* (2016) at p64 in Appx III.

Aside from disagreements over statistics, the volume of pending appeals at the ALJ is another negative effect of the RAC program. As mentioned previously, the appeals process for Medicare FFS claims consists of four administrative levels of review. The first level of appeal is called Redetermination and consists of review of the appealed claim by Medicare policy experts at the MAC. Experts at the QICs review appealed claims at the second level of appeal, which is called Reconsideration. When providers remain dissatisfied, they can appeal the claim to the third level, which is a request to OMHA for hearing before an ALJ.

The total number of appeals filed at Levels 1 through 4 of the process increased significantly between the years 2010 and 2014. The MACs and the QICs are currently meeting their statutory deadlines to process appeals and there is no backlog at the first two levels of appeal. Although they handled fewer claims comparatively, the ALJ Level experienced the largest rate of increase in appeals. The significant growth in volume (more than 1000%) has created a backlog of appeals pending at the third level. By statute, ALJs are ordered to adjudicate appealed claims within 90 days of receipt. However, the average processing time for each third level appeal is currently at more than 1,000 days. Although there was a decrease in the volume of appeals as a result of RAC prohibition of patient status claim review, the processing time has not yet improved, as illustrated in Figures 2 through 4.

87 And a 5th level of judicial review before Federal District Court. CMS. Original Medicare (Fee-for-service) Appeals. Online, available at https://www.cms.gov/Medicare/Appeals-and-Grievances/OrgMedFFSAppeals/index.html
89 Ibid.
90 Ibid.
93 GAO. Report to Congressional Requesters: Medicare Fee-for-Service Opportunities Remain to Improve Appeals Process. (2016) at p11.
95 Ibid.
96 Ibid.
Figure 2: Level 3 (ALJ) Appeals Received\textsuperscript{97}
Source: HHS Agency Annual Reports

Figure 3: Level 3 (ALJ) Appeals Processing Time\textsuperscript{98}
Source: HHS Agency Annual Reports


Although providers are quick to blame the RACs for the increase in appeals, HHS is adamant that RAC-related appeals are just one of several contributing factors. Other drivers of the increase in appeals volume include: higher number of beneficiaries; growth in State Medicaid appeals; and changes and updates to coverage and payment rules. Additionally, according to HHS, the “wide interpretation” of Medicare policy by ALJs led to frequent overturns of claims that were denied by the MAC and the QIC. HHS speculates that providers took notice of this trend and flooded the system with appeals in the hopes for a different – favorable – outcome for their originally denied claims. Regardless of the reasons for the backlog, the Government Accountability Office (GAO) noted in its 2016 Report to Congressional Requesters that it shows no signs of abating because the number of incoming claims continues to surpass the capacity of the ALJs to adjudicate them.

---

99 Ibid.
103 Ibid.
104 GAO. Report to Congressional Requesters: Medicare Fee-for-Service Opportunities Remain to Improve Appeals Process. (2016) at p41.
Providers sought legal relief. The AHA sued HHS’s former Secretary Burwell for lack of timely review of the appeals. The case was initially dismissed, but sent back to District Court on appeal. Ultimately, the Court did not rule as to how it should be accomplished, but did rule that the appeals backlog must be cleared by 2021. The Secretary subsequently appealed and the case was remanded to District Court, yet again in August of 2017. The Court held that District Court must determine that compliance with the task of clearing the backlog is lawfully possible before the Secretary can be required to take action.

III. Improvements Instituted and Their Inadequacies

A. The Attempted Improvements

Changes in the RAC Program

From the start of the RAC program, there have been ADR limits, which dictate the number of ADR requests that RACs can make in a 45-day period. Historically, the ADR limit was 2% of a provider’s claims; however, CMS has since imposed stricter limitations in an effort to reduce provider burden. In late 2015, CMS reduced the ADR limits by three quarters – to 0.5% of a provider’s claims. At the risk of stating the obvious, this means that RACs are not reviewing 99.5% of inpatient hospital claims, a sector known for high rates of improper payments. (Of note, RACs do continue to review inpatient hospital claims, not for admission status, but for incorrectly coded services.) For completeness’ sake, it must also be noted that CMS will change ADR limits based on both RAC and provider performance. Providers who are consistently found to bill accurately will have further reduced ADR limit reductions, and those who demonstrate consistent payment errors will be subject to ADR limit increases. On the other hand, RACs who fail to maintain a 95% accuracy rate in their reviews (as decided by a RAC Validation Contractor) may be subject to progressive ADR limit reductions, among other sanctions.

105 76 F.Supp.3d 43.
106 812 F.3d 183.
108 867 F.3d 160.
109 867 F.3d 160.
110 It is beyond the scope of this paper to detail every change made to the RAC program. Explication of changes will be limited to those pertinent to the issues discussed in this paper.
113 Ibid.
114 Ibid.
115 CMS. Statement of Work (SOW) for the Part A/B Medicare Fee-for-Service Recovery Audit Program – Regions 1-4. at p15.
117 CMS. Statement of Work (SOW) for the Part A/B Medicare Fee-for-Service Recovery Audit Program – Regions 1-4. at p40.
Upon finding an incident of improper payment, RACs must now wait for the duration of the RAC Discussion Period (30 days) before sending the denial to Medicare.\textsuperscript{118} CMS has also altered the timing of the RACs contingency fee payment: RACs do not receive the fee until the claim – if appealed – is adjudicated at the QIC.\textsuperscript{119}

As previously noted, in the fall of 2013, the 2MN Rule was promulgated and RACs were stopped from reviewing patient status claims.\textsuperscript{120} CMS initiated a probe-and-educate program via the MACs as a way to oversee the less stringent reviews of short inpatient hospital stays.\textsuperscript{121} The MACs would deny claims for improper payments, but were instructed to assess provider understanding of the 2MN Rule and offer education to address noncompliance.\textsuperscript{122} While providers preferred this probe-and-educate program to the RAC audit, it came with high administrative costs and MACs were criticized for inconsistent implementation of the program.\textsuperscript{123} Ultimately, in the fall of 2015, CMS gave the authority for inpatient hospital patient status reviews to the Quality Improvement Organizations (QIO) because MACs were not performing to CMS’s satisfaction.\textsuperscript{124} The moratorium on RAC audit of patient status claims was lifted, but the RACs are now only allowed to review observation claims if the QIO identifies a provider with egregious errors and makes a referral to the RAC.\textsuperscript{125}

**Actions to Improve the Appeals Backlog**

Late in 2013, OMHA was struggling to adjudicate the growing numbers of appeals with its comparatively limited human and financial resources.\textsuperscript{126} OMHA even temporarily suspended assignment of appeals to the ALJs.\textsuperscript{127} Although OMHA received supplemental funding to hire additional ALJs, it was not enough to address the growing backlog.\textsuperscript{128} About halfway through 2014, OMHA unsuccessfully tried two new methods to help expedite appeals at the ALJ level: mediation and a statistical sampling initiative.\textsuperscript{129}

In August 2014, HHS acknowledged that it continued to experience extraordinary challenges managing the provider appeals of Medicare overpayment recoveries.\textsuperscript{130} In an effort to effect a meaningful decrease in the volume of pending appeals, CMS offered the option of an settlement to providers.\textsuperscript{131} Any hospital willing to withdraw their pending appeals would receive

\textsuperscript{118} Mary Squire. *RAC: A Program in Distress*, (2015) BYU L. Rev. 219 at p237.
\textsuperscript{119} Id at p238.
\textsuperscript{122} Ibid.
\textsuperscript{123} Ibid.
\textsuperscript{124} Ibid.
\textsuperscript{126} Mary Squire. *RAC: A Program in Distress*, (2015) BYU L. Rev. 219 at p234.
\textsuperscript{127} Ibid.
\textsuperscript{128} Mary Squire. *RAC: A Program in Distress*, (2015) BYU L. Rev. 219 at p239.
\textsuperscript{129} Id at p240.
68% of the net allowable amount of the claims.\textsuperscript{132} The settlement was also intended to ease the administrative burden for all involved.\textsuperscript{133} While the settlement was successful in reducing the number of undecided appeals by an estimated 31%, a large number of pending appeals remained.\textsuperscript{134} Despite these changes, and two additional settlement offers, by 2016, the appeals backlog showed no signs of abating as the number of incoming claims continued to surpass the capacity of the ALJs to adjudicate them.\textsuperscript{135} So, in March of 2017, HHS issued a final rule that introduced other changes.\textsuperscript{136} Decisions made by the Medicare Appeals Council may be designated as precedential in an effort to improve consistency across all level of appeal.\textsuperscript{137} As is true in other, more formal, appellate courts, legal analysis and interpretation of Medicare policy will be binding in future appeals.\textsuperscript{138} However, unlike formal appellate courts, findings of fact will also be binding in future appeals where the relevant facts and evidence are the same.\textsuperscript{139} In theory, this change will increase consistency of decisions across the levels of appeals, reduce the resources spent rendering decisions, and potentially reduce appeals rates.\textsuperscript{140} Attorney adjudicators will be allowed to decide appeals, issue remands to CMS contractors, and dismiss hearing requests if appellants withdraw.\textsuperscript{141} This change would increase the pool of adjudicators allowing appeals to be completed more rapidly.\textsuperscript{142} Other strategies include clarification of regulations; creation of process efficiencies; and addressing other previously identified areas for improvement.\textsuperscript{143}

### Regulatory Changes and the 2MN Rule

The “Improving Access to Medicare Coverage Act of 2013” was a bill introduced in an effort to amend the Social Security Act so that outpatient observation services could be applied toward “qualifying inpatient hospital stay” required for SNF admission.\textsuperscript{144} Ultimately the bill did not make it through the 2013-2014 Congressional session despite the supporters and co-sponsors it garnered.\textsuperscript{145} Providers saw some relief in March of 2013. CMS issued a ruling that allowed hospitals to rebill for outpatient services when contractors denied an inpatient hospital stay during an audit, allowing providers to recoup at least some of the money lost as a result of an overpayment denial.\textsuperscript{146}

\textsuperscript{133} Ibid.
\textsuperscript{134} GAO. Report to Congressional Requesters: Medicare Fee-for-Service Opportunities Remain to Improve Appeals Process. (2016) at p36.
\textsuperscript{135} Id at p41.
\textsuperscript{136} 82 FR 4974.
\textsuperscript{138} 42 CFR §401.109(d)(1).
\textsuperscript{139} 42 CFR §401.109(d)(2).
\textsuperscript{141} Ibid.
\textsuperscript{142} Ibid.
\textsuperscript{143} Ibid.
\textsuperscript{146} Mary Squire. RAC: A Program in Distress, (2015) BYU L. Rev. 219 at p236.
CMS proposed the 2MN Rule in an attempt to clarify the elements that should be present and documented to support an inpatient hospital admission.\(^\text{147}\) The rule was published in the Federal Register in August of 2013, and finalized in the FY 2016 Outpatient Prospective Payment System (“OPPS”) Final Rule, becoming effective as of January 1, 2016.\(^\text{148}\)

When patients present to acute care hospitals, providers have two formal admission choices: outpatient (paid under Part B) or inpatient (paid under Part A).\(^\text{149}\) But there is a third option: to place the patient in the hospital under observation.\(^\text{150}\) The decision between inpatient and observation is where most of the confusion originates.\(^\text{151}\) The Medicare Benefit Policy Manual (MBPM) defines observation as:

“a well-defined set of specific, clinically appropriate services, which include ongoing short-term treatment, assessment, and reassessment before a decision can be made regarding whether patients will require further treatment as hospital inpatients or if they are able to be discharged from the hospital.”\(^\text{152}\)

The expectation per Medicare policy, is that providers can typically make the decision to either discharge or admit the patient after 24 to 48 hours of observation.\(^\text{153}\) This is an outpatient service and billed to Part B.\(^\text{154}\)

An inpatient is defined, quite generally, as a person who “has been admitted to a hospital for bed occupancy for purposes of receiving inpatient hospital services.”\(^\text{155}\) CMS states that providers make this complex medical decision by considering several factors which include: the severity of the patient’s symptoms; the need for, and availability of, diagnostic services; and the medical predictability of a harmful or negative outcome.\(^\text{156}\) Although the MBPM stipulates that admissions are not covered (or noncovered) solely on the length of a hospital stay, providers are given time-based guidance.\(^\text{157}\) Prior to the 2MN Rule, physicians were to use a 24-hour period as a benchmark; in other words, if the patient was expected to require care for 24 hours or more, then inpatient admission was appropriate.\(^\text{158}\) This is clearly at odds with the position that observation can last 48 hours. The 2MN Rule tried to clarify and provide additional specificity. Since its promulgation, the Policy Manual notes that if physicians expect a patient’s care to require a


\(^{148}\) Ibid.

\(^{149}\) CMS. Inpatient or outpatient hospital status affects your costs. Online, available at https://www.medicare.gov/what-medicare-covers/part-a/inpatient-or-outpatient.html

\(^{150}\) Ibid.


\(^{152}\) CMS IOM 100-02 MPBM, Ch6, §20.6(A). (Rev. 215, Issued, 12-18-15, Effective, 01-01-16, Implementation: 01-04-16).

\(^{153}\) Ibid.

\(^{154}\) CMS IOM 100-04 Medicare Claims Processing Manual (MCPM), Ch4, §290. (Rev. 1, 10-03-03, A3-3663, A3-3112.8D, A-01-91).

\(^{155}\) CMS IOM 100-02 MPBM, Ch1, §10. (Rev. 234, Issued: 03-10-17, Effective: 01-01-16, Implementation: 06-12-17).

\(^{156}\) Ibid.

\(^{157}\) Ibid.

\(^{158}\) CMS IOM 100-02 MPBM, Ch1, §10. (Rev. 1, 10-01-03, A3-3101, HO-210).
hospital stay that spans at least two midnights, the benchmark for inpatient admission has been satisfied.\textsuperscript{159}

B. Reasons Why These Changes Are Inadequate

The RAC Program and Observation Reviews

The biggest issue with the changes enacted to the RAC auditing process is the loss of money returned to the Trust Fund.\textsuperscript{160} During the two-year span that the RAC audit of patient status claims was suspended, the RACs were severely limited in their auditing capabilities.\textsuperscript{161} The amount of improper payment recoveries fell from $3.75 billion in 2013 to just $2.39 billion in 2014.\textsuperscript{162} It is estimated that the loss of the audit of observation claims during the moratorium caused the Medicare program to lose more than $8 billion.\textsuperscript{163}

Because the QIO’s mission is to help providers furnish effective, efficient and quality healthcare via education and cooperation with providers without major financial risk, it is generally believed that the QIOs have a better working relationship with providers than RACs do.\textsuperscript{164} However, the American Hospital Association has expressed concerns over the QIO’s inconsistent application of their review process.\textsuperscript{165} Criticism included untimely provision of review results and education which caused 1) missed deadlines for rebilling under Part B, and 2) low improvement rates and higher number of referrals for RAC reviews.\textsuperscript{166} Ultimately, the QIOs struggle with persistent denunciation of patient status claim reviews just as the RACs did.\textsuperscript{167}

QIOs review claims based on the “expectation” and “presumption” review policies outlined in the MBPM after the 2MN Rule:

1. If the provider \textit{expects} a patient to need a hospital stay of at least two midnights, then the inpatient admission is payable under Medicare Part A;

2. The two-midnight \textit{presumption} is that inpatient claims with stays longer than two midnights are appropriate for payment under Medicare Part A.\textsuperscript{168}

\textsuperscript{159} CMS IOM 100-02 MPBM, Ch1, §10. (Rev. 234, Issued: 03-10-17, Effective: 01-01-16, Implementation: 06-12-17).

\textsuperscript{160} CMI. \textit{Medicare Appeals: 2017 Primer}. (2017) at p1.


\textsuperscript{162} Ibid.

\textsuperscript{163} Ibid.


\textsuperscript{165} Id at p10.


\textsuperscript{167} Ibid.

Based on these policies, the QIOs limit their reviews to inpatient claims that did not span two midnights and therefore should have been payable under Part B. Therefore, possible improper payments due to longer, potentially inefficient hospital stays, evade scrutiny.

Finally, the QIO program runs at a much higher cost than the RAC program. The RAC program literally pays for itself: the net savings to the Trust Fund are reported after all the costs of the program, including contingency fees, are accounted for. Funding for the QIO program is defined as mandatory (not discretionary) spending and is not subject to the appropriations process. Running the QIO program actually drains the Medicare Trust Fund further since its costs are financed directly from it. Even when RACs were at their busiest, the costs to run the program were consistently less than costs to run the QIO program (see Figure 5).

**Figure 5: QIO vs RAC Program Costs**

Sources: HHS. Agency Annual Reports and CMS. Annual Report to Congress: The Administration, Cost, and Impact of the QIO Program for Medicare Beneficiaries for Fiscal Years

<table>
<thead>
<tr>
<th>Year</th>
<th>RAC Program</th>
<th>QIO Program</th>
</tr>
</thead>
<tbody>
<tr>
<td>FY 2011</td>
<td>$128.5</td>
<td>$308</td>
</tr>
<tr>
<td>FY 2012</td>
<td>$228.1</td>
<td>$373</td>
</tr>
<tr>
<td>FY 2013</td>
<td>$454.1</td>
<td>$512</td>
</tr>
<tr>
<td>FY 2014</td>
<td>$460.9</td>
<td>$579</td>
</tr>
<tr>
<td>FY 2015</td>
<td>$95.9</td>
<td>$668</td>
</tr>
<tr>
<td>FY 2016</td>
<td>$59.9</td>
<td>$776</td>
</tr>
</tbody>
</table>

---

169 Ibid.
172 CMS. *Annual Report to Congress: The Administration, Cost, and Impact of the QIO Program for Medicare Beneficiaries for Fiscal Year 2016*. At p.5.
173 CMS. *Annual Report to Congress: The Administration, Cost, and Impact of the QIO Program for Medicare Beneficiaries for Fiscal Year 2016*. At p.5.
Continued Issues with Appeals at the ALJ Level

It is thought that, instead of reducing the appeals burden, the settlement offered by CMS in 2014 could have inadvertently provided incentive for more appeals instead of less, since providers might hope for yet another settlement from the continued appeals backlog.\textsuperscript{175} More importantly, though, is what the OIG characterizes as the “wide interpretation” of Medicare Policy at the ALJ level.\textsuperscript{176}

Appellants who are dissatisfied with the QIC’s Reconsideration decision can file a request for hearing along with supporting documentation to OMHA who assigns the appeal to an ALJ.\textsuperscript{177} The ALJs perform de novo reviews of the claims and the relevant Medicare policies and documentation.\textsuperscript{178} In other words, the ALJs make independent determinations for the claims at issue and are not bound by prior decisions or findings.\textsuperscript{179} ALJs are bound by statutes, regulations, NCDs, and CMS rulings.\textsuperscript{180} ALJs are directed to pay substantial deference to LCDs and CMS manual guidance, however, they can, and do, choose to decline to follow the guidance in those documents.\textsuperscript{181}

The data collected by CMS and OMHA is not sufficient to substantiate the extent to which ALJs decline to follow LCDs and CMS manual guidance.\textsuperscript{182} Nevertheless, HHS and the OIG have noted that the ALJs’ disinclination to apply LCDs as they are applied at the lower levels of appeal is a problem.\textsuperscript{183} OMHA went on to designate this lack of understanding and compliance with LCDs and manual guidance as a key issue for improvement.\textsuperscript{184} Additionally, there is substantial subjectivity in the ALJs’ application of policy to the facts of each appeal to the extent that, examination of many ALJ decisions from hearings that involved an LCD shows that the policy was applied differently than at the lower level of appeal.\textsuperscript{185}

It has been reported that ALJs tend to render favorable decisions for providers based on the intent of Medicare policy being met, rather than the “letter of the law” being satisfied.\textsuperscript{186} This subjectivity and inconsistent application of Medicare policies is thought to be a major contributing factor to the overturn rate of lower-level denials which (as noted previously) hovers near 50\%.\textsuperscript{187}

\textsuperscript{175} CMI. 2016 State of the RAC Program. Online, at p9, accessed at \url{http://medicareintegrity.org/wp-content/uploads/2016/03/2015StateOfTheProgram-FINAL.pdf}
\textsuperscript{176} CMI. 2016 State of the RAC Program. Online, at p9, accessed at \url{http://medicareintegrity.org/wp-content/uploads/2016/03/2015StateOfTheProgram-FINAL.pdf}
\textsuperscript{177} GAO. Report to Congressional Requesters: Medicare Fee-for-Service Opportunities Remain to Improve Appeals Process. (2016) at p8.
\textsuperscript{178} Ibid.
\textsuperscript{179} Ibid.
\textsuperscript{180} Ibid.
\textsuperscript{181} GAO. Report to Congressional Requesters: Medicare Fee-for-Service Opportunities Remain to Improve Appeals Process. (2016) at p9.
\textsuperscript{182} GAO. Report to Congressional Requesters: Medicare Fee-for-Service Opportunities Remain to Improve Appeals Process. (2016) at 23.
\textsuperscript{183} Ibid.
\textsuperscript{184} Ibid.
\textsuperscript{185} GAO. Report to Congressional Requesters: Medicare Fee-for-Service Opportunities Remain to Improve Appeals Process. (2016) at 23.
\textsuperscript{187} Id at p79.
The high overturn rate reflects poorly on the expertise and competency of the ALJs.188 ALJs do not have the authority to base decisions on their own standards, and this loose interpretation of the regulations undermines the integrity of the Medicare program.189

The 2MN Rule & Other Regulatory Efforts

While CMS has explained that the 2MN Rule was meant to respond to calls for improvement from all stakeholders involved in patient status reviews, there is a “unanimity of dislike” of the rule among all of those same stakeholders.190 Hospitals have generally opposed the rule, arguing that it is administratively onerous, overly complicated, and that it fails to support physician judgment.191 The rule constitutes a distinct regulatory shift from clinical criteria to time-based criteria – which seems to contradict CMS’s own language in the MBPM that states admissions are not covered based on length of stay alone.192 Providers contend that the rule disregards the level of care necessary for safe patient treatment and undermines the “complex medical judgment” that CMS indicates physicians must apply when making the admission decision.193

The 2MN Rule has done little-to-nothing to update or revise the three-day statutory requirement for SNF coverage.194 In fact, several lawsuits have been filed in an attempt to eliminate observation based on the theory that it improperly denies beneficiaries of rehabilitation coverage upon discharge from an acute care hospital.195 Further, the rule did little to ameliorate the higher costs of copayments faced by beneficiaries when they are treated under observation.196 If an inpatient admission is denied for medical necessity, hospitals can rebill the claim under Part B.197 The beneficiary will be responsible for the subsequent deductibles and copays applicable under part B.198 The amounts beneficiaries are liable for under Part B often exceed the amounts they would be responsible for under Part A.199

CMS contractors have been instructed to focus their review efforts on hospital stays that do not cross two midnights since they would not be presumed appropriate for Part A payment per the 2MN Rule.200 However, there remains a paucity of regulatory guidance concerning what

188 Id at p80.
189 Id at p81.
193 Id at p14.
194 Id at p16.
198 Ibid.
199 Ibid.
constitutes inpatient or observation once the requisite two-midnight stay is satisfied. The presumption of reasonableness provides little protection for providers in the end, since their judgment will still be open to scrutiny and admissions still vulnerable to denial.

CMS’s regulatory solutions have focused, historically, and under the 2MN Rule, on redefining or clarifying inpatient status and making it more distinct from outpatient or observation. The persistent exclusive consideration of these dichotomous admission options only aggravates the difficulty of a decision that clearly requires mitigation.

IV. Suggested Revisions

A. Define and Codify Observation as an Admission Status

Observation has existed since the 1960s, but it is still inconsistently used and improperly billed because of poor definitions and poor regulatory guidance. Pundits maintain that misuse of observation actually worsened after both the initiation of the RAC program and after the promulgation of the 2MN Rule.

The use of observation has legitimate benefits, and a viable proposal for regulatory reform should not simply eliminate it. Care provided in observation beds imparts cost-effective clinical flexibility to physicians when patients present with 1) conditions that are not truly appropriate for admission, but are unstable, or uncertain and potentially serious enough to warrant close monitoring, or 2) when a diagnosis is known, but the clinical course is unpredictable, or 3) when deciding where to place the patient for care is difficult.

“Outpatient” and “inpatient” are already defined and codified in Federal Regulations. Despite the existence of these definitions, there is overlap between services rendered in each setting, perpetuating confusion. Additionally, the financial ramifications are pronounced due to the drastic difference in reimbursement based on outpatient versus inpatient status. That difference in reimbursement becomes difficult to substantiate when a patient receives the same services to treat the same conditions, regardless of the label on the admission order form.

203 Id at p16.
204 Ibid.
205 Ibid.
206 See 42 CFR §§410.2 and 440.2.
209 8 St. Louis U. J. Health L. & Pol’y 147 at p15.
210 Id at p16.
211 Id at p16.
Meanwhile, the concept of observation remains in limbo, floating somewhere between outpatient and inpatient care.\textsuperscript{213} CMS should use already-established guidance (its own, and from physician specialty expert groups like the American College of Emergency Physicians (ACEP)\textsuperscript{214}) to define and outline proper usage of observation. Observation can then be codified as an intermediary patient status, with its own separate payment. This would bridge the gap between outpatient and inpatient, clinically and financially.

B. Restore the Audit of Patient Status Claims to the RACs

Safeguarding Medicare and the Trust Fund by reducing improper payments is vital and of greater importance than ever before.\textsuperscript{215} The authorization to contract with private entities is codified in the SSA.\textsuperscript{216} The 2MN Rule is clear that inpatient hospital claims will still be evaluated by medical review contractors to ensure the medical necessity of services provided.\textsuperscript{217} While ‘protecting the integrity of the Trust Fund’ is part of CMS’s definition of the core functions of the QIOs, the QIOs are more focused on patient safety and quality of care.\textsuperscript{218} The RACs, on the other hand, are specifically tasked with finding and recovering improper payments.\textsuperscript{219} Given that the FFS improper payment rate remains unacceptably high\textsuperscript{220} displacing the RACs from patient status reviews does not align with Medicare’s expectation that the RACs focus on the identification of improper payments having the greatest impact on the Trust Fund.\textsuperscript{221}


\textsuperscript{216} 42 USC § 1395kk.


\textsuperscript{219} CMS. Statement of Work (SOW) for the Part A/B Medicare Fee-for-Service Recovery Audit Program – Regions 1-4. at p1.


C. Codify the Discussion Period as Part of the RAC Review Process

CMS requires RACs to allow for a Discussion Period, as noted previously. After the RAC denies a claim, providers are given 30 days to submit additional medical records to support the billing of that claim. Further, RAC must allow for the option of a verbal “peer-to-peer” discussion between the RAC medical director and provider’s physician representative if requested. However, providers are not required to participate in this portion of the process. Therefore, providers are not affording themselves of a potentially valuable opportunity to obtain a favorable outcome without becoming mired in the formal Medicare Appeals Process (MAS). If more providers exploited the Discussion Period, there is great potential for fewer claims being formally appealed, avoiding the administrative and financial burden of going through the formal MAS.

The Discussion Period can remain entirely separate from the Medicare Appeals Process and be codified with language that is similar to that in the CFR for the MACs:

“In conducting a redetermination, the contractor (MAC) reviews the evidence and findings upon which the initial determination was based, and any additional evidence the parties submit or the contractor obtains on its own.”

And, although the Discussion Period is separate from the formal appeal process, and would remain so, codifying it is consistent with HHS’s actions to provide more timely adjudication of claims. HHS’s actions include: efforts aimed at reducing the number claims at the first two levels of appeal; and efforts aimed at resolving the backlog of appeals at the third level of appeal. A mandatory RAC Discussion period would contribute to these efforts.

D. Make LCDs Mandatory Authority, Rather Than Persuasive Authority

The framework around NCDs and LCDs parallels that of Federal and State law. The echo of the principle of Federalism can be heard in nature of coverage determinations. Federal Laws apply to the entire nation, and State Laws apply to their respective states. Similarly, NCDs, which are developed by CMS, apply to the entire nation, and LCDs, which are developed by the MACs, apply to each MAC’s jurisdiction.

---

222 CMS. Statement of Work (SOW) for the Part A/B Medicare Fee-for-Service Recovery Audit Program – Regions 1-4, at p30.
223 Ibid.
224 Ibid.
225 Ibid.
227 CFR § 405.948.
228 GAO. Report to Congressional Requesters: Medicare Fee-for-Service Opportunities Remain to Improve Appeals Process. (2016) at p32.
229 Ibid.
Because of the plenary authority given to the states in the Constitution, state legislatures have greater ability to regulate across a range of issues than does Congress. The number of state statutes regarding public health, for example, is far greater than those passed by Congress. So, too, do LCDs, outnumber NCDs. Further, LCDs offer much greater specificity than NCDs. NCDs describe the circumstances required for services to be covered by Medicare, nationwide – the minimum coverage requirement. LCDs stipulate the clinical circumstances under which services are considered to be reasonable and necessary.

Federal Law has precedence over State Law. NCDs have primacy over LCDs: LCDs must be consistent with all statutes, regulations, Medicare rulings and national coverage, payment, and coding policies (emphasis added).

The issue of mandatory versus persuasive authority is where the similarities diverge. Although Federal Law preempts State Law, on issues of State Law, the state’s highest court retains mandatory authority for all other courts – even Federal courts. NCDs, like Federal laws, are binding on all Medicare carriers and contractors and are also binding on ALJs during the appeals process. Not so with LCDs. A claim reviewed by a contractor in a particular MAC’s jurisdiction must comply with that MAC’s applicable LCDs. Although ALJs are directed to give substantial deference to LCDs, they are not binding.

Generally, courts look to agencies as having the greatest expertise regarding the relevant issues, yet this is not always true in the relationship between ALJs and the MACs. LCDs are developed after experts at the MACs consider medical literature, the advice of local medical consultants and societies, and comments from both the public and the provider community. LCDs might be issued when there is no NCD, or to further define an existing NCD. They can be established when frequent claim denials are received (or anticipated) after post-payment claims.
reviews; or when there is verification of a widespread problem involving high-dollar services. And despite all of these facts, ALJs are not required to follow their guidance.

As previously noted, the substantial subjectivity and inconsistency seen in the application of LCDs by ALJs has been recognized as a key issue for improvement by OMHA. Making the LCDs binding would be consistent with OMHA’s opinion, and would help to improve the consistency of the ALJs’ decisions.

V. Conclusion

While there are many who fear that healthcare spending is unsustainable and will lead to the insolvency of the Medicare Trust Fund, there are also skeptics who dismiss the notion as popular myth. It is important, though, to keep in mind that Medicare is the single largest payer for healthcare services in the country and is vital to the country’s healthcare system. Due to its sheer size and scope, most providers would not generate enough revenue to remain in business without the beneficiaries and subsequent reimbursement for their healthcare that would be lost if the Medicare Program should fail. Both the general public and CMS, consequently, have a vested interest in safeguarding the Trust Fund by ensuring payment accuracy. The vast majority of improper payments are honest mistakes due to administrative and documentation errors. The lack of specific guidance leads both providers and contractors, to err when making the proper admission determinations. Claims must be reviewed to promote payment accuracy. The typically competing interests of the stakeholders involved are currently in agreement that the regulations, as they stand, do not suffice. Medicare policy and regulations must be modified so that providers, RACs, and ALJs, alike, can make accurate decisions that align with coverage and payment rules.

---

249 GAO. Report to Congressional Requesters: Medicare Fee-for-Service Opportunities Remain to Improve Appeals Process. (2016) at p23.
252 Ibid.
253 Ibid.