



# Comprehensive Application of Predictive Modeling to Reduce Overpayments in Medicare and Medicaid

*Prepared by:*

**The Lewin Group, Inc.**

*June 25, 2009*

*Revised July 22, 2009*

## Table of Contents

<b>Background .....</b>	<b>1</b>
<b>Pre-Pay Predictive Modeling .....</b>	<b>2</b>
Walk-Through of Pre-Pay Predictive Modeling Process.....	4
Examples: Cases Where Predictive Modeling Identified Improper Claims Not Identified by Traditional Rule-Based Screens and Edits .....	5
<b>Estimates of Potential Savings .....</b>	<b>7</b>
<b>Conclusion.....</b>	<b>9</b>

## Background

Improper payments for health care goods and services are estimated to be in the range of 3% to upwards of 10% of total health care expenditures nationally.<sup>1</sup> Improper overpayments for fee-for-service medical claims in Medicare and Medicaid are estimated by the Center for Medicare & Medicaid Services (CMS) to be on the order of \$10.4 billion for Medicare and \$12 billion (federal share) and \$21 billion (total computable) for Medicaid in FY 2007.<sup>2</sup> These estimates suggest error rates of 3.6% for fee-for-service Medicare claims and 8.3% for Medicaid fee-for-service claims.

Current methods used by CMS to reduce improper payments in Medicare fee-for-service include a limited application of pre-payment screening, editing and selective review of claims, conducted by the Medicare Administrative Contractors (MAC). However, most resources are devoted to post-pay review activities.

CMS operates several fraud and abuse programs that partner with law enforcement agencies to audit claims and providers, identify potential fraud and recoup overpayments to providers. In 2006, CMS' Medicaid Integrity Group established the role of Medicaid Integrity Contractors (MIC), whose purpose is to review and educate Medicaid providers as well as audit claims submitted by providers and identify overpayment of funds. The Medicare-Medicaid Data Matching Project (Medi-Medi) began in 2001 and, as of 2007, is operating in 10 states. The program attempts to identify fraud and abuse patterns across both programs that would not necessarily be evident in reviewing the programs individually. The Deficit Reduction Act of 2005 appropriated \$12 million for FY 2006, \$24 million in FY 2007, \$36 million in FY 2008, \$48 million in FY 2009 and allows for this program to be funded at \$60 million annually in FY 2010 and beyond. Since the program's inception, approximately "50 Medi-Medi cases have been referred to law enforcement, \$15 million in overpayments have been referred for collection, and \$25 million in improper payments have been caught before erroneous payments were made."<sup>3</sup>

Additionally, CMS is currently in the process of consolidating the work of existing Program Safeguard Contractors (PSC) and Medicare Drug Integrity Contractors (MEDIC) to form Zone Program Integrity Contractors (ZPIC), which will have the responsibility to ensure the program integrity for all Medicare claims, under Parts A, B, C and D as well as coordinate with the Medi-Medi program as appropriate.<sup>4</sup>

---

<sup>1</sup> See, for example, Health Care Anti-Fraud Association. (2009); *The problem of health care fraud*; Federal Bureau of Investigation. (2007); *Financial crimes report to the public: Fiscal year 2007*; PricewaterhouseCoopers' Health Research Institute (2008); *The price of excess: Identifying waste in healthcare spending*. (This report also included assessments of annual excess costs in operational processes (e.g., claims processing, ineffective use of information technology, paper prescriptions, consumer behavior, and clinical services).)

<sup>2</sup> These estimates are taken from the FY 2007 Comprehensive Error Rate testing (CERT) program for Medicare and the Payment Error Rate measurement (PERM) program for Medicaid.

<sup>3</sup> *Medicare Health Care Fraud & Abuse Efforts: Hearing before the Committee on the Budget, U.S. House of Representatives*, (2007) (testimony by Timothy Hill, Chief Financial Office, Center for Medicare and Medicaid Services).

<sup>4</sup> Center for Medicare & Medicaid Services. (2008). *CMS enhances program integrity efforts to fight fraud, waste and abuse in Medicare*. Retrieved from <http://www.cms.hhs.gov/apps/media/press/release.asp?Counter=3291>

Prior to the formation of the ZPICs, PSCs were primarily responsible for identifying potential fraudulent activities, referring instances of potential fraud to law enforcement and conducting proactive data analysis to identify potential fraud in Medicare Part A and Part B. According to an OIG review of PSCs published in July 2007, PSCs reported to CMS that \$54,673,571 had been identified in connection with PSC investigations and \$119,053,255 was reported in connection with referrals to law enforcement in 2005. However, the same OIG report identified a lack of consistency across PSCs in terms of production and found that there was no correlation between the budget of each PSC, the oversight responsibility and the number of new investigation or law enforcement referrals. Further, the OIG found that proactive data analysis, a primary PSC function, produced minimal new investigations or case referrals in 2005.<sup>5</sup> In addition, PSCs and now ZPICs largely have retrospective focus and have a limited role in developing or implementing preventative fraud detection measures.

Finally, CMS also began a three-year pilot in 2005 for the Medicare Recovery Audit Contractor (RAC) program, which initially operated in three states and then was expanded to six states in 2007. The RAC contractors are responsible for identifying and recouping Medicare overpayments and are paid based on the percentage of improper payments corrected. The RACs collected \$992.7 million in overpayments to providers and corrected \$37.8 million in underpayments to providers, as of March 27, 2008, representing 0.3% of total Medicare claims received during the same time period.<sup>6</sup> CMS is in the process of implementing the RAC program permanently.

## Pre-Pay Predictive Modeling

The predictive modeling method considered here is applied prior to the payment of a claim. If successful, this has the obvious advantage over post-pay methods in that an improper payment is prevented from being made in the first place.

A comprehensive pre-pay system consists of an initial tier of rule based screens and edits. The next is a predictive model that identifies improper payments, fraud and abuse by “scoring” the claim, based on its characteristics. Finally, for issues that cannot be addressed pre-pay, post-pay review and analysis can audit, identify and recover funds that may slip through the pre-pay methods, and otherwise be lost to waste, fraud and abuse. A comprehensive pre-pay system, which includes predictive modeling, can be significantly more effective than relying largely on traditional post-pay “pay and chase” methods.

The claims processing systems for Medicare at the federal level and Medicaid at the state level include pre-pay “screens and edits”. These screens and edits can produce automated denials, automated corrections, and flags of specific claims for manual review. These pre-pay tools, while valuable, are based on very simple rule-based logic. These types of simple “screens and edits” would remain in place and even be improved. The predictive modeling approach supplements the simple logic of these screening and editing methods, providing a more powerful method to detect claims that have a high probability of payment error.

---

<sup>5</sup> Department of Health and Human Services Office of Inspector General. (2007). *Medicare’s Program Safeguard Contractors: Activities to detect and deter fraud and abuse.*

<sup>6</sup> Center for Medicare & Medicaid Services. (2008). *The Medicare Recovery Audit Contractor (RAC) program: An evaluation of the 3-Year demonstration.*

As claims enter the system, and pass through traditional pre-pay screens and edits, each claim would be “scored” for its risk of improper payment.<sup>7</sup> The method of scoring is based on proven relationships between claim characteristics, provider characteristics, and risk of overpayment. Claims with relatively high scores would have a high risk that if the claim were paid as it was entered into the system, it would result in an overpayment.

The scoring model is built from a set of provider characteristics that are highly correlated with fraudulent or inappropriate claims billing. These characteristics that have been incorporated in commercial applications include unbundling, upcoding, percent of time a provider exceeds 8-hour workdays, recent increase in weekends worked, recent increase in lines billed per patient across multiple claims, recent increase in modifiers per claim line per service, among others. Statistical methods, guided by expert understanding of billing methods, are used to find the underlying set of characteristics that provide the explainable sources of variations in the data. With this statistical method, each underlying characteristic contributes to the score assigned to each claim line. Line scores are then added, and if the total score exceeds a threshold, the scoring model flags the entire claim for further investigation. When appropriate, medical records are requested, then reviewed by both clinical and fraud investigators. From the results of the medical records review as well as analysis conducted by modelers, the model scoring is further refined to reduce the false positive rates. This is the closed feedback loop that helps improve the model precision.<sup>8</sup>

The predictive model employs advanced methods to detect fraudulent patterns across claims by considering multiple factors that are too subtle and complex for traditional rules-based screens and edits to identify. The patterns are often more complex than any single rule or multiple set of rules.

Historically, pre-payment screening methods have seen limited application. The primary reason for this is that the methods generate a large number of “false positives.” These are claims that scored relatively high on the risk scale yet, upon manual investigation, are found to be correct. This false positive rate raises costs, because the claims incorrectly identified must be reviewed. In addition, it increases the cycle time, or payment period, for those providers.

Recent experience in the commercial sector indicates that the predictive model methods have largely mitigated these problems through improved accuracy and methods, including integration of manual review in the process.<sup>9</sup> The predictive accuracy of the new model is much greater, with applications in the commercial sector achieving accuracy rates in excess of 80%.<sup>10</sup> The accuracy rate after manual review increases to well in excess of 90%. Moreover, evidence from the commercial sector suggests that less than 5% of denials under the system are appealed and overturned.

False positive are significantly fewer using these methods, reducing the number of costly manual reviews that produce no finding of error. Moreover, procedures have been developed for provider “self-audit” so that, once high risk claims are identified through the predictive modeling, providers are offered the opportunity to adjust or withdraw their claim on-line. Only

---

<sup>7</sup> The concept is not unlike the scoring method the Internal Revenue Service uses to select tax returns for audit.

<sup>8</sup> These scoring models can use a wide range of underlying statistical and analytical techniques. A discussion of these specific techniques and the ones used in commercial applications is beyond the scope of this paper.

<sup>9</sup> See, for example, Ingenix. (2009). *Payment integrity: Harnessing the power of technology to detect and deter fraudulent, erroneous and abusive claims*, for a brief report of actual experience with a commercial payer.

<sup>10</sup> Higher accuracy rates can be achieved when higher “cutoff” scores are applied to screen claims. Accuracy rates in excess of 80% are obtained for cutoff score used in the commercial applications.

after this period for “voluntary correction” is the process of manual review begun for claims that have not been satisfactorily modified or withdrawn by the provider. This further reduces the costs associated with manual review.

## Walk-Through of Pre-Pay Predictive Modeling Process

The diagram below, Figure 1, illustrates how the predictive modeling process is incorporated into the claims processing system.

Electronic claims enter the system, and are subject to the types of screens and edits that are currently in place (or, perhaps, improved screens and edits.)<sup>11</sup>

Those claims that pass through these initial screens and edits are “scored” based on their characteristics and the predictive model scoring equation.<sup>12</sup>

Based on the “score” the claim receives (which is related to the probability that the claim is in error or fraudulent) the claim is then assigned to pay, deny, or “pend” categories. Denied and pended claims are manually reviewed, within a day’s time, before being sent back to the provider. Those that are “pended” are flagged for further review and analysis.

Based on the score and the characteristics of the pended claim and the provider, the flagged claim is sent to “self-audit”, where the provider has a chance to review the claim on-line, and modify or withdraw the claim.

- If the claim is withdrawn or appropriately adjusted, no further review action is taken.
- If the provider makes no change to the claim status, the claim is briefly reviewed and sent either for full manual review or is paid

Other flagged claims, based on their score and characteristics, are sent to staff review and are examined by expert clinicians, investigators, and clinical coders who often obtained medical records from the provider for these “pended” claims.<sup>13</sup>

Based on the manual review, the claim is either paid or denied.

A key decision in the process is setting the “cutoff” score for flagging the claim. The higher the score, the higher is the probability that the claim is in error. A very high cutoff score means that fewer claims will be flagged, but the accuracy of those flagged will be very high. There will be few “false positive” errors. At the cutoff scores set in commercial applications, very few of the claims are pended based on the scoring algorithm and the cutoff score.<sup>14</sup> At this level, the accuracy rate is greater than 80%. That is, there are fewer than 20 “false positives” for every 100 claims flagged. Moreover, though a small proportion of claims are

---

<sup>11</sup> In one example of a commercial system, claims pass through over 9 million edits.

<sup>12</sup> A predictive model is **estimated** using historical data and a predictive model scoring equation is produced from this estimation process. That equation is then **applied** to new data to score each new record in the dataset.

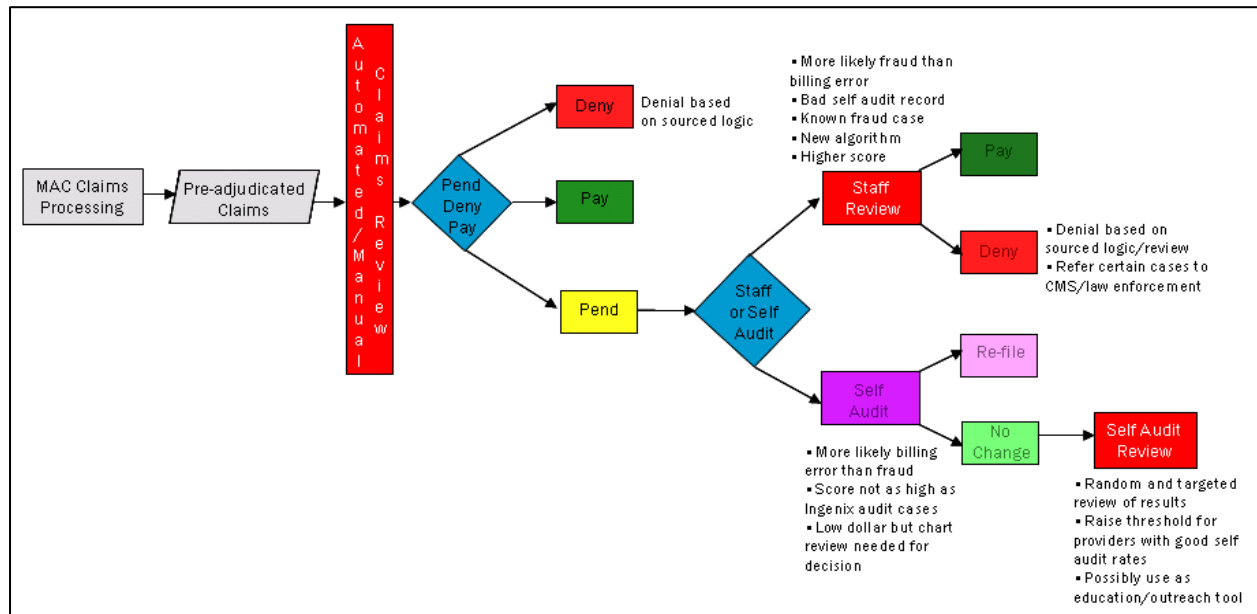
<sup>13</sup> The Ingenix prospective review predictive modeling solution pends between 0.2 %and 0.5% of paid claims.

<sup>14</sup> See the note above.

flagged at this level of the cutoff score, experience in the commercial sector indicates that about 2.6% of total payments are saved.<sup>15</sup>

The system is one of continual evaluation and learning. Algorithms are adjusted regularly based on the results of the manual reviews, analysis of false positives, and other information. Hence, in principle, the system is both improving over time, and adjusting to the behavior of providers.

Figure 1: Process Incorporating Pre-Pay Predictive Modeling



### Examples: Cases Where Predictive Modeling Identified Improper Claims Not Identified by Traditional Rule-Based Screens and Edits

The examples below consist of a number of actual cases where predictive modeling was applied in the commercial market. In these cases, the claim was flagged based on the relatively high “score” given to it by the predictive model. It would have not been flagged or adjusted using traditional logical screens and edits.

In each case, the service was selected by the predictive model as a service with an elevated probability of being in error because of scores it received on several predictor variables or factors. The scores are based on a combination of variables or characteristics associated with the claim and the provider. The claim was flagged because of the high total score it received. It would be difficult, if not impossible, to have identified this claim using the definitive logic of rule-based screens or edits. The claims in this example, in fact, passed through a panoply of front-end screens and edits without being flagged prior to being scored by the predictive model, and were flagged based on the predictive model score. Subsequent review confirmed that the

<sup>15</sup> Note that this result is based on the way the algorithm is configured. Higher scores are provided to large dollar valued claims, compared to smaller claims, for the same probability of error.

claim was in error. The examples, then, illustrate, the value predictive modeling adds above rule-based screens and edits.<sup>16</sup>

### **Claim 1**

- This claim, submitted by an orthopedic surgeon, was selected for medical record review based on the “score” it received, which was driven by several factors: 1) historically high frequencies of general modifier use compared to peers; 2) specific modifier usage, and 3) performance of unusual procedures within the specialty.
- Medical record review revealed a clear narrative describing shoulder joint arthritis. A simple aspiration of shoulder joint fluid and injection of lidocaine and steroid was described, without description of imaging guidance.
- **The following was supported by record:**
  - Evaluation and management service
  - Drainage / injection of joint/bursa
- **The following was billed, but unsupported by record:**
  - Fine-needle aspiration under imaging guidance
  - Intravenous injection of lidocaine
  - Injection of 80 mg of methylprednisolone

---

### **Claim 2**

- This claim, submitted by a chiropractor, was selected for medical record review on the basis of the “score” it received. This score was driven by the following factors: 1) historically high frequencies indicating a high concentration of one service by that provider on the same day; 2) mismatches of diagnosis and procedure codes; and 3) specific modifier usage.
- Medical record review documented chiropractic manipulation and the application of neurostimulator.
- **The following was supported by record:**
  - Chiropractic manipulation
  - Application of neurostimulator
- **The following was billed, but unsupported by record:**
  - Evaluation and management services that include detailed history and medical decision-making of moderate complexity
  - Therapeutic exercises (separately identifiable from chiropractic manipulation)
  - Manual therapy (separately identifiable from chiropractic manipulation)

---

### **Claim 3**

- This claim, submitted by an orthopedic surgeon, was selected for medical record review on the basis of the high “score” it received. This score was driven by: 1) historically high frequencies of general modifier by that provider compared to peers; and 2) performance of high-level (upcoded) evaluation and management services within the specialty.
- Medical record review revealed 8 check-marks on a list, as documentation of the history and physical examination; brief documentation of 3 injections “dexa / lido 3x3”. Patient’s presenting problem is one word: “Triggers.” Four diagnoses are documented only as four abbreviated single words: “migr, insom, fatig, fibr”

---

<sup>16</sup>Note once again, that predictive modeling complements traditional pre-pay methods. It will be applied while traditional screens and edits (though, perhaps, significantly improved) remaining place.

- The following was supported by record:
  - Trigger point injections in more than 2 muscles
- The following was billed, but unsupported by record:
  - Evaluation and management services that include detailed history and medical decision-making of moderate complexity
  - Separately identifiable therapeutic injection

These examples illustrate the added value of predictive modeling, above edits and rules, and demonstrate that predictive models that are accurate can detect potential improper claims before they are paid.

## Estimates of Potential Savings

This proposal would introduce pre-pay predictive scoring and review of high risk fee-for-service claims.

Estimates of the potential savings from introducing this comprehensive are shown in Table 1.<sup>17</sup>

**Table 1. Original Estimates of Potential Savings from Pre-pay Predictive Model**

Billions (\$)	2010	2011	2012	2013	2014	2010-2014	2010-2019
Medicare/Medicaid Programs	0	6.2	10.0	10.8	11.6	38.6	113

To test the reasonableness of these estimates, we conducted an analysis of the potential savings from the comprehensive application of pre-pay predictive modeling to Medicare and Medicaid fee for service. To do this, we made several assumptions.

1. The first, and most important, is the net rate of savings to the programs. There is some empirical evidence on this rate. In a commercial application, it was found that 2.6% of total fee-for-service professional service payments were saved.<sup>18</sup> This savings rate estimate is in addition to the savings achieved by programs in place prior to the implementation of the pre-pay predictive model process. To be conservative, our estimate of the savings rate applied in the table below is 1.5%.<sup>19</sup>
2. To further assess the reasonableness of the estimate, we apply the most recent published fee-for-service error rate for Medicare and Medicaid. The FY 2007 payment error rate for

<sup>17</sup> UnitedHealth Center for Health Reform and Modernization. (2009). *Health Care Cost Containment: How Technology Can Cut Red Tape*. (Working Paper 2.)

<sup>18</sup> Memorandum from Simon Rosenstein to Andrew Asher (December 2008).

<sup>19</sup> Empirical evidence on the savings rates, evaluated under controlled conditions, is available for professional payments (physicians and other health professionals). Subject matter experts believe that the savings rate will vary by service type. For example, it is believed that the savings rate in durable medical equipment may be about 6% on average, while the savings rate for inpatient, institutional claims may average about 1.25%. Based on the subject matter experts' estimates of how savings will vary across service type, an overall savings rate of 2.6%, the rate actually observed for professional payments, is also about the average rate expected across all programs. The 1.5% rate applied here is a conservative estimate of this average rate, given the estimates of subject matter experts and the empirical evidence on professional services programs. Moreover, the conservative rate will more than adjust for program costs, while remaining conservative.

Medicare was 3.6% and for Medicaid 8.3%.<sup>20</sup> For this analysis, we assume that these error rates are constant over the project period. Arguably, the error rates should decline.

3. For Medicaid and Medicare expenditures, we are using the CMS Office of the Actuary National Health Expenditure Projections, 2008-2018, as the basis for program growth. We assume that the proportion of the total programs, as measured by expenditures, to which the pre-pay predictive modeling solution will be applied, remains constant over time. We assume that the federal share for Medicaid remains constant at 57%.
4. We are not including CHIP in this analysis.
5. We assume that other payment integrity programs in effect now continue over time at the same levels as currently. The estimated savings shown below are in addition to those that would be achieved by existing pre-pay and post-pay programs.
6. We assume partial implementation in 2011, with full implementation in 2012.

Under the assumptions presented immediately above, Table 2 presents our estimates.

**Table 2. Estimates of Potential Savings from Pre-Pay Predictive Model (\$ Billions)**

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2010-2019
<b>Medicare</b>											
Medicare Expenditures	515.5	547.4	584.9	627.9	674.4	729.1	789.4	857.7	931.9	1012.5	
Subject to Pre-pay Model	344.5	365.9	390.9	419.7	450.7	487.3	527.6	573.3	622.9	676.7	
Total Estimated Errors	12.4	13.2	14.1	15.1	16.2	17.5	19.0	20.6	22.4	24.4	
Estimate of Savings	0.0	2.7	5.9	6.3	6.8	7.3	7.9	8.6	9.3	10.2	65.0
<b>Medicaid</b>											
Medicaid Expenditures	419.1	452.0	487.5	527.4	571.8	621.3	675.6	735.2	800.7	872.0	
Subject to Pre-pay Model	321.8	347.1	374.4	405.0	439.1	477.1	518.8	564.6	614.9	669.7	
Total Estimated Errors	26.8	28.9	31.2	33.7	36.6	39.7	43.2	47.0	51.2	55.8	
Federal Errors	15.3	16.5	17.8	19.2	20.8	22.7	24.6	26.8	29.2	31.8	
Estimate of Savings (total)	0	2.6	5.6	6.1	6.6	7.2	7.8	8.5	9.2	10.0	63.6
Estimate of Savings (Federal)	0	1.5	3.2	3.5	3.8	4.1	4.4	4.8	5.3	5.7	36.2
<b>Combined Medicare and Medicaid</b>											
Estimate of Total Savings	0.0	5.3	11.5	12.4	13.3	14.5	15.7	17.1	18.6	20.2	128.6
Estimate of Total Federal Savings	0.0	4.2	9.1	9.8	10.5	11.4	12.4	13.4	14.6	15.9	101.2

These estimates suggest that total federal savings from implementing pre-pay predictive modeling could be about \$9.1 billion in FY 2012, if fully implemented by the beginning of that

<sup>20</sup> Some observers believe that these error rates are an underestimate of the true error rate.

year, and \$101.2 billion over the period of FY 2011-FY 2019. Total savings, that includes the state's share of Medicaid, would be \$128.6 billion over the period FY 2011-FY 2019.

Our estimated savings rate, 1.5% is a conservative estimate taken from commercial sector experience. The lower rate will presumably account for savings that may not occur in other programs designed to reduce improper payments because of the effectiveness of pre-pay predictive modeling and for program costs, while remaining a conservative estimate. However, while total savings from all other programs are difficult to document precisely, it appears that CMS sponsored programs save under \$1 billion per year.<sup>21</sup> Hence, it appears that even if savings from other programs were reduced by an additional 50% as a result of pre-pay methods, the net savings would remain quite substantial.

Our estimates in Table 2 are in the same range as the estimates of the savings provided in Table 1.<sup>22</sup> Our estimate of federal savings in FY 2012, the first full year of implementation, of \$9.1 billion is close to the Table 1 estimate of \$10.0 billion in combined Medicare and Medicaid savings. The estimates of total federal savings over the period 2011-2019 are within about 12% of each other.<sup>23</sup>

## Conclusion

Our estimates suggest that the original estimates of the potential savings from this program, from Table 1, are reasonable. A key assumption underlying our independent estimates in Table 2 is that the savings rate experienced in the commercial sector can be projected to the federal sector. To be conservative, we reduced that estimate by almost 40%. This conservative estimate also mitigates the risk that other programs aimed at reducing improper payments may grow, possibly reducing the potential savings from any single program. Moreover, even if savings from existing fraud and abuse programs were reduced, the net savings from the pre-pay predictive modeling appears to remain significant.

In addition, we assume that the underlying error rate does not, itself, decline over time. This, in a sense, is counterfactual in that preventing errors prior to payment will result in a reduction in the error rate. However, reduction from this source will not affect the estimated savings, because the pre-pay program itself must be there to prevent the errors. Over time, however, one might expect the pre-pay predictive model to serve as a deterrent. In that case, directly measured savings may decline somewhat but, arguably, this "deterrent effect" should be considered a benefit when assessing the value of the program.

Finally, one additional point that is not mentioned above. While Medicare claims processing has standard formats and policies applied nationally under relatively homogenous processing systems, there is significant variation in state Medicaid claims formats, policies and processing. This may make the Medicaid pre-pay predictive model somewhat more costly to implement, compared to Medicare.

---

<sup>21</sup> See the *Background* section of this paper for a brief summary of published savings estimates.

<sup>22</sup> Tables 1 estimates are from UnitedHealth Center for Health Reform and Modernization. (2009). *Health Care Cost Containment: How Technology Can Cut Red Tape*. (Working Paper 2.)

<sup>23</sup> Moreover, our estimates in Table 2 include only the potential savings from Parts A and B of Medicare. If we were to include Part D, estimated savings over the period 2011-2019 would increase to within about 1% of the Table 1 estimate for the same period, or a total of about \$112 billion.