An Integrated Decision-making Framework for Medical Audit Sampling

Tahir Ekin
McCoy College of Business, Texas State University
t_e18@txstate.edu

Abstract

The loss of three to ten percent of annual health care expenditures to fraudulent transactions makes medical audits paramount. In order to handle the size and complexity of medical claims, the use of analytical methods and information technology tools to aid in medical audits is necessary. In general, sampling frameworks are utilized to choose representative claims. However, these are not integrated within audit decision-making procedures. As a novelty, this paper presents an integrated decision-making framework for medical audit sampling. We propose a simple but effective optimization method that uses sampling output and enables auditors address the trade-offs between audit costs and expected overpayment recovery. We use U.S. Medicare Part B claims payment data to demonstrate the framework.

1. Introduction

Health care spending has an upward trend worldwide, especially in developed countries. The increasing median age of the population coupled with costs make this a serious concern for policy makers. In the United States, annual health care expenditures reached $3.2 trillion or $9,990 per person in 2015, which accounted for 17.8 percent of the nation's Gross Domestic Product [1]. It is reported that up to ten percent of this spending is lost to medical overpayments in the form of fraud, waste and abuse [2]. These overpayments impact both the government and tax-payers with direct cost implications. They also diminish the ability of medical systems to provide quality care to beneficiaries.

The levels and types of these health care overpayments vary. The most common instances are seen through identity fraud, improper coding and kickback payment schemes. For instance, identity fraud include cases where a provider uses the identification of patients for illegitimate billing. Whereas improper coding corresponds to incorrect billings where the billed procedures do not match the actual procedure provided. Such overpayments can take place as a result of clerical mistakes or deliberate attempts to increase revenue. Some examples of improper coding are listed as upcoding, unbundling, multiple billing and phantom billing. Upcoding refers to billing for a more expensive service or procedure than the one actually performed. Unbundling is submitting separate claims for services or supplies that should have been grouped together. Providers billing for the same claim more than once and billing for procedures that were never provided are instances of so called multiple and phantom (ghost) billing. More sophisticated overpayment schemes include fraudulent networks that are based on kickback payment schemes and self-referrals.

Medical audits correspond to the manual investigation of medical claims by domain experts to determine their legitimacy. These audits are generally costly and time-consuming. The size of the health care system prohibits all submitted claims to be audited, and this results in the requirement of sampling. Sampling refers to choosing a representative subset of claims of interest. Then, these audit results are projected to the population. The differences within claims makes sampling a challenging process, hence various types of sampling are utilized. These include but are not limited to simple random sampling, stratified and multi-stage sampling methods. Overall, medical audit decisions involve consideration of audit costs, expected recovery amounts, accuracy of the extrapolations (overpayment estimation) as well as the regret cost due to incorrect outcomes. In particular, the audit costs include the investigation time spent by the experts and the physical resources. The recovery costs include immediate recovery due to audits as well as the projected recovery. Overall, even a small improvement in the audit sampling decisions while satisfying the governmental guidelines is crucial.

The use of statistics, optimization and information technology has become an integral part of many health

HICSS

care processes. Medical overpayments and fraud assessment are not exceptions. However, to the best of our knowledge, there are not any decision models in medical audit sampling literature. In order to fill this gap, this paper proposes an integrated medical audit sampling decision analytics framework. In doing so, the proposed optimization model uses the output of an information theoretic stratified sampling method, and considers initial and additional audit costs and expected recovery amount as well as a budget constraint. The proposed integrated decision-making framework has the potential to improve the decision-making for medical audits while ensuring the statistical validity. While the sampling method is not the emphasis, we present an analysis and optimization of medical audit sampling with an emphasis on trade-offs between audit costs and expected overpayment recovery. This provides a semi-automated decision-making alternative within the health care fraud assessment systems. Second, as a minor contribution to the general audit sampling literature; the proposed framework is general in that it can be used in other audit settings.

The paper is organized as follows. The following section presents the current practice and literature related to the medical audit sampling. Section 3 outlines the utilized sampling methods whereas Section 4 describes the proposed optimization model. Section 5 presents the utilized Medicare Part B claims data and illustrates the use of the method with an analysis. The paper concludes with an overview and a discussion of future research directions in Section 6.

2. Medical Audit Sampling

In the U.S., governmental medical services are mainly provided through the federal and state programs of *Medicare* and *Medicaid* which are administered by *The Centers for Medicare & Medicaid Services* (CMS). Governmental organizations have launched many initiatives to oversee the health care spending and decrease overpayments [3]. All the payments to a provider cannot be audited; thus, the government relies on statistical sampling to choose a set of claims for audits. Then they extrapolate from sample to population. This requires the appropriate choice of medical claims sample that is followed by the overpayment investigations. Will Yancey [4] provides a comprehensive list concerning these legal sampling procedures and the parties involved in U.S. governmental medical insurance programs. A review of the literature can be found in Ekin et al. (2018) [5], while Woodard (2015) [6] demonstrates the use of sampling by U.S. governmental programs to reveal overpayments. The proportion of overpaid claims and the overpayment amount in a given population are particularly of interest. Medical auditors can use *Rat-Stats* [7], which is a statistical software package offered by the *Office of Inspector General, Office of Audit Services*, to assist with the statistical analysis. *Rat-Stats* can perform three main functions: determination of sample size, generating random numbers to select the sample and providing inference. This tool can be used to conduct both attribute and variable appraisal, see *Rat-Stats Manual* [8] for an overview.

Simple random sampling has been widely used in medical audits since it is easier to understand, perform and communicate with others. However, it may not be the method of choice when the auditor has knowledge about population subgroups. As an alternative, stratified sampling is based on separating the population into mutually exclusive, homogeneous segments (strata) by a stratification variable, and then drawing samples from each segment (stratum). The objective is to choose strata in a way that minimizes the within-group differences while maximizing the between-group differences. Separate estimates of overpayment are made for each stratum and the weighted stratum estimates yield an overall projected overpayment. There are three major benefits in using stratified sampling [8]. First, it can be more efficient since it can provide a smaller margin of error for the same sample size due to smaller weighted sum of the strata variances, especially when the measurements within strata are homogeneous [9]. Secondly, it can provide additional information about each stratum. This can be beneficial when the auditor needs more precise estimates for a specific group or wants to ensure that a particular group is represented in the sample. Thirdly, stratified sampling gives flexibility to employ different sampling methods within each stratum for cost or precision considerations. Overall, it can be argued that stratification can be the method of choice if the auditor has sufficient information or interest in particular subgroups. However, it should be noted that the decision maker needs to pay attention to the sampling design which includes the choice of number of strata, the stratum boundaries, and the sample sizes of each stratum.

The only multi-stage medical audit sampling approaches in literature are by Ignatova and Edwards (2008) [10] and Musal and Ekin (in-press) [11]. Ignatova and Edwards [10] suggest using a pilot (probe) sample first, and checking if the number of overpaid claims are higher than a threshold. In case of small overpayment levels in the first stage, the

medical investigation can be halted; otherwise, the population can be investigated further. Musal and Ekin [11] propose an iterative stratified sampling method that uses Lindley's entropy measure to evaluate the expected amount of information. A practical alternative is to combine single stage sampling methods within a multi-stage audit framework, for which *Rat-Stats* can be utilized. For instance, one can first obtain a sample from primary variables such as hospitals proportional to their size. Then, the sample can be further partitioned with respect to a secondary variable such as provider type. Multistage sampling can be a beneficial and efficient alternative, especially when the evaluation of the tradeoff between precision and cost is important. Hence, this paper utilizes a modified version of the iterative information theoretic stratified sampling approach of [11].

The population of interest in these sampling procedures are usually the payment amounts to a provider, of some of which consist of overpayments. A payment amount associated with a claim can result in one of three outcomes when audited. A claim can be classified as completely legitimate, completely illegitimate or partially overpaid. Current governmental sampling guidelines [9] recommend to use the lower limit of a one sided 90 percent confidence interval for the total overpayments as the recovery amount from the provider under investigation. Using the lower bound allows for a reasonable and fair recovery without requiring a tight precision to support the point estimate, sample mean. In other words, the state is protected from recovering an amount greater than the true value of erroneous payments. The simple expansion, ratio and regression-based estimators are among the widely used estimation methods in audits; see [12] for an overview. The simple expansion method is based on computing the mean overpayment in the sample, and multiplying it by the population size, whereas the ratio estimator for the total overpayment is computed as the product of the total payment value and the ratio of the sample overpayment and sample payment values. Medical claims data is well known to exhibit skewness and non-normal behavior, thus, requiring large sample sizes for accurate estimation [13]. Alternative approaches include but are not limited to the minimum-sum method of Edwards et al. (2003) [14], the zero-one inflated mixture model of Ekin et al (2015) [13] and the Bayesian mixture model of Musal and Ekin (2017) [15].

Despite all these developments in medical audit sampling; in literature and practice there are not any decision frameworks for medical audit sampling. Even a simple decision analysis setup can be useful for formal utilization of the sampling output. This paper aims to propose a simple but effective optimization model to fill this gap.

3. Information theoretic stratified sampling

This section introduces the utilized information theoretic stratified sampling approach. It is a modified version of the iterative multi-stage method of [11] which utilizes Neyman Allocation [16] for initial allocation of samples to strata and Lindley's entropy measure for additional allocation. The expected recovery for each allocation decision alternative will be computed as an input for the proposed decision-making framework.

First, we introduce the following notation. The total number of strata is L. The total number of claims in the population is $N = \sum_{h=1}^{L} N_h$ where N_h is the number of claims in stratum h. The unknown total number of overpayments in the population is denoted as $K = \sum_{h=1}^{L} K_h$ where K_h is the number of overpaid claims in stratum h. The payment and overpayment amounts of claims in stratum h are represented by the vectors $\mathbf{X_h} = \{\mathbf{X_{(h,1)}}, ..., \mathbf{X_{(h,N_h)}}\}$ and $\mathbf{Y_h} = \{\mathbf{Y_{(h,1)}}, ..., \mathbf{Y_{(h,N_h)}}\}$, respectively. We assume that $\mathbf{X_h}$ is known but $\mathbf{Y_h}$ is only known after investigation. $\boldsymbol{\rho}$ denotes the proportion of overpaid claims, a vector of size L, that consists of ρ_h, the proportion of overpaid claims in each stratum h.

The symbol k denotes the number of overpaid claims in a sample of size n. Once the samples are allocated to strata, the sample payment and overpayment amounts of claims in stratum h are represented by the vectors $\mathbf{x_h} = \{\mathbf{x_{(h,1)}}, ..., \mathbf{x_{(h,n_h)}}\}$ and $\mathbf{y_h} = \{\mathbf{y_{(h,1)}}, ..., \mathbf{y_{(h,n_h)}}\}$, respectively. The number of overpaid claims in stratum h is denoted as k_h in a sample of n_h.

3.1. Data pre-processing and initial allocation

Data pre-processing includes the determination of number of strata, L and the strata boundaries. Then, the claims and the respective payment values are allocated to each stratum with respect to chosen stratum boundaries. This needs to be followed by the step of determining the sample sizes of each stratum for each decision alternative for the initial number of medical claims to be investigated. This influences the precision and cost of a stratified sampling design. Neyman Allocation allows the auditor to consider the variance of the estimates while assuming the cost to sample from these strata to be same. Hence, it is utilized in this paper for initial allocation.

In particular, for initial investigation, $n^{(init)}$ number of units are allocated to strata based on the standard deviation, σ_{X_h} and $\bar{X}_h$, mean payment of each stratum,

$h = 1, ..., L$. The sample size for stratum h can be written as:

$$n_h^{(init)} = \frac{n^{(init)} N_h \sigma_h}{\sum\limits_{h=1}^{h=L} N_h \sigma_h} \quad (1)$$

where

$$\sigma_{X_h}^2 = \frac{\sum\limits_{i=1}^{i=N_h} (X_{h,i} - \bar{X}_h)^2}{N_h - 1}$$

Next, we discuss how to do the additional iterative allocation using Lindley's entropy measure.

3.2. Additional allocation

The approach by [11] uses Lindley's entropy measure in a Bayesian framework to quantify the expected information content of sampling from each stratum. We use their idea to determine the additional allocation within the stratified sampling setup. In particular, the steps of the utilized sampling framework are:

1. For stratum h, obtain $p(\rho_h | n_h^{(add)}, k_h^{(add)})$, which is the posterior distribution of the proportion of overpaid claims conditional on the draw of the additional samples, $n_h^{(add)}$ and outcome if that is legitimate or overpaid, $k_h^{(add)}$:

   The probability distribution of k is Binomial with parameters (n, ρ) and the prior distribution of ρ is defined via Beta distribution with the hyper parameter vector $\omega = (\alpha, \beta)$. Due to conjugacy, this leads to the posterior distribution of ρ as:

   $$p(\rho | k, n) \sim Beta(\alpha + k, \beta + (n - k)). \quad (2)$$

2. For stratum h, obtain the distribution $p(K_h | k_h^{(add)})$:

   The posterior distribution of the number of overpaid claims in the population, in a particular stratum K_h, can be evaluated by using Hyper-Geometric distribution. Due to the conjugacy of Beta-Binomial prior probability and Hyper-Geometric likelihood, the posterior probability $p(K_h | k_h)$ follows a Beta-Binomial distribution [17].

3. For stratum h, obtain $p(k_h^{(add)} | n_h^{(add)})$ and compute $E[\Delta_h(k_h^{(add)})]$, the expected information gain from the additional sample(s) $k_h^{(add)}$:

   The expected information gain of an additional sample, $n_h^{(add)}$ from stratum h, with the outcome $k_h^{(add)}$, is denoted as $E[\Delta_h](k_h^{(add)})$. We use $k_h^{(add)}$ and k_{ff} to refer to the number of overpaid claims, k, from future draws of claims, $n_h^{(add)}$ and n_{ff} respectively, given the data and updated parameters. $E[\Delta_h](k_h^{(add)})$ can be written as a double expectation over $k_h^{(add)}$ and k_{ff}:

   $$fE[\Delta_h(k_h^{(add)})] = E[log \frac{p(k_{ff} | k, k_h^{(add)}, n_h^{(add)})}{p(k_{ff} | k)}] \quad (3)$$

4. Determine h^*, the stratum with the highest expected information gain and sample from h^*:

   $$h^* = argmax_h E[\Delta_h(k_h^{(add)})] \quad (4)$$

5. Estimate the overpayment recovery amounts for the all strata as well as the overall population:

   In line with the governmental guidelines, we aim to be conservative and prevent incorrect recovery demands from the audited provider. Therefore, the 10th percentile of total number of overpaid claims, $K_{h,0.1}$ is retrieved from the posterior distribution of K. It is multiplied with the mean payment to obtain the expected total overpayment recovery for stratum h, $Y_{rec,h}$ as

   $$Y_{rec,h} = K_{h,0.1} \bar{X}_h \quad (5)$$

   The aggregate total recovery amount, Y_{rec} can be found via:

   $$Y_{rec} = \sum\limits_{h=1}^{h=L} Y_{rec,h}. \quad (6)$$

Our emphasis in this paper is not on the details of the sampling method. Please see [11] for background of the information theoretic sampling approach and

the derivation of related probability distributions. It should be noted that our sampling setup differs from [11] in that our interest is on recovery amount, instead of the estimation errors. In this paper, we have only provided a brief presentation to demonstrate the sampling framework. In particular, our focus is on the computation of the expected recovery amount for decision alternatives of initial investigation size and additional number of samples to be collected. These will be inputted to the following integrated decision-making framework. It should be noted that following optimization model is general in which it can be used with the output of any iterative sampling framework.

4. Decision-making Framework

We present an optimization model to make audit sampling decisions. This decision-making framework enables auditors to consider the trade-offs between audit costs and expected recovery while choosing representative medical claims within the budget. In the following, we present the decision variables, parameters and the related functions within the optimization model.

Decision Variables
$n^{(init)}$: initial number of medical claims to be investigated and allocated via Neyman Allocation
$n^{(add)}$: additional number of medical claims to be investigated and allocated via information theoretic approach

Parameters
r: recoupment percentage
$\bar{y}$: mean overpayment
Y_{rec}: expected recovery amount
c_1: unit audit cost for initial investigations
c_2: unit audit cost for additional investigations
B: total audit budget

The optimization model is written as:

max

$$rY_{rec} + (n^{(init)} + n^{(add)})\bar{y} - c_1 n_h^{(init)} - c_2 n_h^{(add)}$$

subject to

$$c_1 n^{(init)} + c_2 n^{(add)} \leq B, (n^{(init)}, n^{(add)}) \in A$$

where Y_{rec} can be computed via Equation (6), and it is an indirect function of decision alternatives. This can be written as $Y_{rec}(n^{(init)}, n^{(add)})$ and A is the set of decision alternatives.

The objective is to minimize the total audit cost and maximize the expected recovery gain. The decisions to be made are the initial and additional number of claims to sample; $n^{(init)}$ and $n^{(add)}$.

In particular, the recovery consists of two functions; recovery due to audits and expected recovery from the population. The recovery due to audits is deterministic in the sense that on average, after the audits, overpayment amount $(n^{(init)} + n^{(add)})\bar{y}$ is demanded from the provider under investigation.

In addition, with respect to governmental guidelines, the sample overpayment can be extrapolated to the population. Such expected overpayment recovery is denoted as Y_{rec} and can be computed for each decision alternative as aforementioned. The government is only able to recoup a certain percentage of the inferred population overpayments. Therefore, the recovery gain from the population is discounted by using a recoupment percentage, r; so it becomes rY_{rec}.

The total audit cost consists of initial and additional sample allocation cost. It is assumed that additional samples cost more than the initial samples, since that would correspond to extra investigation after the initial setup. Therefore $c_1 < c_2$, and these costs are $[c_1 n^{(init)}$ and $c_2 n^{(add)}]$, respectively. Lastly, total cost cannot exceed the budget, B.

5. Illustration

This paper uses the publicly available *2008 CMS Outpatient Procedures* data file [18]. The medical claims and billing process have relatively remained same, so the use of a ten year old data set has no impact on the validity of the results. A variety of procedure codes that had frequent overpaid billings in investigations [19] are chosen as candidates for medical audits. The procedure codes of interest are *J9265, J9310, J0475* and *J9041*, which correspond to injections in surgeries. The resulting data set consists of 8278 claims with payment values.

Stratification is conducted with respect to payment values. Our assumption is that payment values are correlated to the overpayment values. The actual allocation of payments to $L = 5$ strata is done using the R package *GA4Stratification* [20] so that each stratum consists of increasing dollar amounts of payments. Table 1 presents the descriptive statistics of payment values.

The boundaries of 5 strata are determined as $\{(0 - 375), [375 - 1600), [1600 - 2650), [2650 - 3550), [3550 - 4301]\}$. This is how payments are allocated to each stratum and $\mathbf{X_h} = \{\mathbf{X_{(1,h)}}, ..., \mathbf{X_{(1,N_h)}}\}$ for $h = 1, ..., 5$ is determined.

Stratum	Mean	Sd	Median	N_s
s=1	69.40	46.92	40.00	3949
s=2	915.69	180.87	70.00	1402
s=3	2335.37	244.61	2500.00	588
s=4	3076.60	206.76	3000.00	1675
s=5	4012.65	246.71	4100.00	664
Overall	1298.47	1441.22	600.00	N=8278

Table 1. Descriptive statistics of payment

Next, let's illustrate the proposed framework which is based on Neyman allocation and information theoretic sampling. Table 2 presents the initial and additional number of investigated claims with allocation to strata for select set of decision alternatives. For instance, in the third line there are 45 claims that are allocated in the initial investigation, and 10 claims to be allocated additionally. For each decision alternative, the iterative sampling framework is used to allocate samples and draw the claims as overpaid or legitimate. The number of overpaid claims in each stratum in Table 2 represents the output of one simulation, and aims to help demonstrate how the sampling framework works. For instance, second stratum has 17 claims of which two are overpaid. The decision alternative of $n^{(init)}, n^{(add)} = (55, 0)$ in the fourth line also corresponds to a total of 55 claims that are all allocated at the initial phase. The allocation to strata can be recognized to be different then $n^{(init)}, n^{(add)} = (45, 10)$ despite the same total size of audited claims. The output of optimization model determines the optimal allocation by assessing if the expected recovery would justify the audit costs. The proposed optimization model is relatively simple, and can be solved using Excel Solver in a second.

$n^{(init),(add)}$	s=1	s=2	s=3	s=4	s=5
45,0	8 (0)	10 (1)	6 (2)	14 (11)	7 (7)
45,5	8 (0)	12 (1)	9 (5)	14 (11)	7 (7)
45,10	8(0)	17 (2)	9 (5)	14 (11)	7(7)
55,0	9 (0)	13 (1)	7 (2)	18 (11)	8 (7)

Table 2. Allocation to each stratum for select decision alternatives

For illustration, we have defined the discrete set of decision alternatives, A in the ranges of $(30, 90)$ for $n^{(init)}$ and $(10, 30)$ for $n^{(add)}$. We have assumed the unit sampling costs to be c_1=400; c_2=1,000 respectively. These costs include but not limited to the investigation setup, salary and expenses of the auditor. The recoupment percentage, r is assumed to be 0.1, using the information in the *Annual Report to Congress on the Medicare and Medicaid integrity programs* for fiscal years 2013 and 2014 [21]. In

particular, 490 billion dollars were spent on Medicare in 2014. Assuming 10 percent overpayment rate, out of 49 billion dollars overpayments, only 4.765 billion dollars were recovered. This approximately corresponds to a recovery rate of 10 cents for each overpaid dollar. The budget, B is set as $500,000$. All parameters can be modified depending on the nature of the medical audit.

Table 3 presents the sample overpayment amounts and expected recovery amounts for a select set of decision alternatives. Overall, we have found the optimal decision that gives the maximum total revenue as $(n^{(init)} = 45, n^{(add)} = 30)$.

$n^{(init)}$	$n^{(add)}$	$\bar{y}$	Y_{rec}
45	10	840.12	550.34
45	20	900.91	600.11
45	30	979.51	612.30
60	10	930.23	527.91
60	20	910.02	548.23
60	30	920.16	590.12
90	10	904.76	570.43
90	20	886.17	537.23
90	30	924.80	610.23

Table 3. Sample overpayment and expected recovery amounts for select decision alternatives

The proposed framework enables the auditor to assess the trade-offs between costs and expected recovery. The expected recovery, which is an output of the sampling framework, is a function of decision alternatives. Therefore, additional sampling can be preferred despite its higher unit cost when the expected recovery becomes higher due to learning as part of the information theoretic sampling.

Figure 1 presents a snapshot of these relationships for select decision alternatives. Audit costs are in the x axis, whereas the expected recovery is given in the y axis. The optimal decision of $n^{(init)} = 45, n^{(add)} = 30$ reveals that higher cost for additional samples is worth for the gain in expected recovery.

The relative advantage of the information theoretic method is smallest in the cases of small additional sample sizes. The monetary gain contribution is shown to decrease with increasing total sample size [11]. It should be noted that since auditors generally utilize relatively small samples for medical audits, even small improvements in sampling designs is crucial.

6. Conclusion

The extent of health care fraud is in billions of dollars. The size and heterogeneity of medical claims data require the use of analytical and information
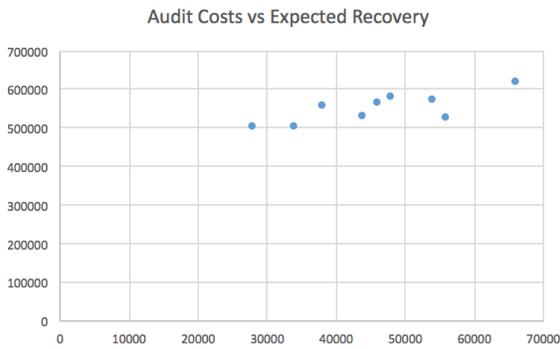
Figure 1. Audit costs vs expected recovery for select decision alternatives

technology methods to aid in medical audits. Although a various number of sampling methods are utilized for audits, there are not any decision models for medical audit sampling. This paper fills that gap by presenting an integrated medical audit sampling decision analytics framework. The proposed method enables auditors address the trade-offs between expected recovery and cost while having valid overpayment amount estimates. U.S. Medicare Part B claims payment data is used to demonstrate an illustration of the proposed decision model and discuss trade-offs.

In particular, the proposed method builds on a modified version of the information theoretic stratified sampling framework of [11] and extends it by incorporating its output within a simple but novel optimization model. A demonstration of the trade-offs between the cost and expected recovery for different decision alternatives of initial and additional investigation sizes is provided. In general, the auditors should assess at what point the additional costs are not justified by the expected recovery.

There are a number of directions for future research that can address the limitations of our work. It should be noted that the optimal decisions of the proposed optimization model are sensitive to the choice of the parameter values. A comprehensive sensitivity analysis among the ratio of the unit costs and recoupment percentage can help auditors assess the trade-offs better. The current setup includes utilizing the sampling framework and computing the recovery amounts for the decision alternatives of consideration. This can be extended into a fully integrated multi-stage decision model which can be run sequentially. The initial investigation decision at the first epoch can impact the potential decisions at later stages. A more comprehensive approach that incorporates the distributions of number of overpaid claims and recovery amount, can be considered as an alternative. The

current setup is based on using the output of the statistical model, not the entire distribution. Using the distributions within a full Bayesian approach can provide additional insights.

On a related note, the proposed optimization model does not consider the potential regret cost due to overpayment estimation errors. As expected, the literature shows that estimation errors decrease with larger sample sizes. Such incorporation of regret cost can be feasible for cases with known and established overpayment patterns.

Analytical methods and optimization along with good use of information technology allow health care organizations to combine, integrate, secure and analyze large quantities of data. The debate about use of automated systems has recently gained more steam with the help of the recent computational advances. However, we should note that the complexity and the nature of the health systems coupled with the legal requirements may make semi-automated systems a more realistic option. We believe the proposed framework is a good alternative in that direction.

References

[1] CMS, "NHE Fact Sheet." The Centers for Medicare & Medicaid Services. https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/nationalhealthexpenddata/nhe-fact-sheet.html, 2017. Accessed: 07/01/2017.

[2] H. Shin, H. Park, J. Lee, and W. Jhee, "A scoring model to detect abusive billing patterns in health insurance claims," *Expert Systems with Applications*, vol. 39, no. 8, pp. 7441–7450, 2012.

[3] OIG, "Annual report for fiscal year 2013." http://oig.hhs.gov/publications/docs/hcfac/FY2013-hcfac.pdf, 2013. Accessed: 10/23/2017.

[4] W. Yancey, "Sampling for Medicare and other claims." http://www.willyancey.com/sampling-claims.html, 2012. Accessed: 1/18/2018.

[5] T. Ekin, F. Ieva, F. Ruggeri, and R. Soyer, "Statistical methods for medical fraud assessment: Exposition to an emerging field," *International Statistical Review. OnlineFirst. doi:10.1111/insr.12269*, 2018.

[6] B. Woodard, "Fighting healthcare fraud with statistics," *Significance*, vol. 12, no. 3, pp. 22–25, 2015.

[7] "Ratstats - statistical software." http://oig.hhs.gov/compliance/rat-stats/index.asp, 2016. Accessed: 04/30/2018.

[8] "Ratstats user guide." http://oig.hhs.gov/organization/oas/ratstats/UserGuide2010_04js.pdf, 2016. Accessed: 04/30/2018.

[9] CMS, "Program memorandum carriers transmittal b-01-01." http://www.cms.gov/

`Regulations-and-Guidance/Guidance/`
`Transmittals/downloads/B0101.pdf`, 2001.
Accessed: 03/03/2018.

[10] I. Ignatova and D. Edwards, "Probe samples and the minimum sum method for Medicare fraud investigations," *Health Services and Outcomes Research Methodology*, vol. 8, no. 4, pp. 209–221, 2008.

[11] R. Musal and T. Ekin, "Information theoretic multi-stage sampling framework for medical audits," *Applied Stochastic Models in Business and Industry. OnlineFirst. doi:10.1002/asmb.2338*, 2018.

[12] D. Guthrie *et al.*, "Statistical models and analysis in auditing: Panel on nonstandard mixtures of distributions," *Statistical Science*, vol. 4, pp. 2–33, 1989.

[13] T. Ekin, R. M. Musal, and L. V. Fulton, "Overpayment models for medical audits: multiple scenarios," *Journal of Applied Statistics*, vol. 42, no. 11, pp. 2391–2405, 2015.

[14] D. Edwards, G. Ward-Besser, J. Lasecki, B. Parker, K. Wieduwilt, F. Wu, and P. Moorhead, "The minimum sum method: a distribution-free sampling procedure for medicare fraud investigations," *Health Services and Outcomes Research Methodology*, vol. 4, no. 4, pp. 241–263, 2003.

[15] R. M. Musal and T. Ekin, "Medical overpayment estimation: A Bayesian approach," *Statistical Modelling*, vol. 17, no. 3, pp. 196–222, 2017.

[16] J. Neyman, "On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection," *Journal of the Royal Statistical Society*, vol. 97, no. 4, pp. 558–625, 1934.

[17] D. L. Grosh, "A Bayes sampling allocation scheme for stratified finite populations with Hyperbinomial prior distributions," *Technometrics*, vol. 14, no. 3, pp. 599–612, 1972.

[18] CMS, "Basic Stand Alone (BSA) Medicare claims public use files (PUFs)." `https://www.cms.gov/bsapufs`, 2010. Accessed: 11/07/2016.

[19] N. Youngstrom, "Medical-necessity audits gain steam, hit on chemo, cardiac; watch for new lcds." `https://bit.ly/2sVy6KH`, 2015. Accessed: 11/07/2017.

[20] T. Keskintürk and Ş. Er, "A genetic algorithm approach to determine stratum boundaries and sample sizes of each stratum in stratified sampling," *Computational Statistics & Data Analysis*, vol. 52, no. 1, pp. 53–67, 2007.

[21] CMS, "Annual Report to Congress on the Medicare and Medicaid integrity programs for fiscal years 2013 and 2014." The Centers for Medicare & Medicaid Services. `https://www.cms.gov/outreach-and-education/outreach/partnerships/fraudpreventiontoolkit.html`, 2016.