Calculating Financial Accuracy Rates in Health Insurance Claims Audits

Lisa A. Gardner, Ph.D.
Drake University

ABSTRACT

This article examines varying financial accuracy rate calculations, and, when working with right-skewed health claims amounts distributions, varying approaches to sampling and choices of stratifying variables. After considering the strengths and limitations of common approaches, consideration is given to the best way to sample and stratify health insurance claims. The methods advocated for herein, when taken in combination, represent a departure from common practices and standards, including the vaguely phrased 'random sampling' approach required by the AICPA and others.

Subject: Accounting, Insurance
Article Type: Editorially-Reviewed Journal Article

INTRODUCTION

As healthcare costs continue to outpace general inflation rates, employers and government insuring entities like Medicare and Medicaid increasingly rely on health claims audits to help rein in costs (Hurlock, 2010) and manage fraud (Mcguffie, 2011) (State News Service, 2016). Audits of dependent eligibility, and of clinical, disability, electronic prescription drug, medical, and prescription drug claims can yield short-term savings by facilitating recoveries of overpayments made in error.[1] Identifying overpayments saves the employer or government money, but identifying underpayments, too, helps ensure that the plan operates with integrity.[2] While this should be a goal of any plan administrator, paying claims correctly also helps ensure that employees and government insurance beneficiaries have more favorable attitudes towards their employers (Wells, 2010), government, or towards hospital staff (Bohnsack and Hawig, 2012). However, claims that are not paid correctly often require employee or beneficiary intervention, which can result in dissatisfaction.[3]

Cost savings from audits may vary from a few hundred thousand dollars (i.e., see (Lubbers, 2010) and (Ramunni, 2016)), to millions (i.e., see (Benedict, 2015), (Kalish, 2013), (Cherf, 2010) or (Stegman, 2008)). Figures like these give reason enough to consider an audit. Other reasons why employers and government insurers might want audits of third party administrators supporting self-funded plans, or, of health insurance companies, include to meet compliance obligations due to the fiduciary responsibilities of ERISA, SSAE-16, and Sarbanes-Oxley 404[4]; to establish a performance baseline for a new plan administrator or to evaluate a plan administrator in relation to his or her peers using claims payments metrics; to provide trustees and staff with information helpful in negotiating future contract performance requirements and fees (Claim Technologies Incorporated, 2008) (Cusick and Anderson, 2014); and to consider some of the impact of selected major policy changes (e.g., requirements to cover adult children up to age 26).

Auditors make several decisions when evaluating health care claims, including how to measure a plan's financial accuracy, and which claims ought to be reviewed using what means. Their decisions may be shaped by AICPA guidelines and by state or federal laws and regulations.

The AICPA does not prescribe financial accuracy rate calculations, leaving them largely to the discretion of auditors. However, the AICPA is not silent on the question of how to select claims for auditing, supporting simple random sampling.

Auditors of Medicare or Medicaid, carried out by private firms called Recovery Audit Contractors (RACs) who have no affiliations with insurance intermediaries, health insurance carrier or

---

[1] Overpayments may be caused by care that is not medically necessary; documentation that is missing, incomplete or incorrect; duplicate claims payments; and inaccurate coding.

[2] Causes of underpayments include errors in applying coinsurances or co-payments; incorrectly denying charges as duplicate payments; incorrectly denying eligible expenses; PPO calculation errors; and UCR calculation errors.

[3] Audits can also help prevent future payment errors by identifying and correcting existing processing or documentation errors, many but not all of which may be systematic and thus repeating. Some attention may be paid to claims turnaround times – the rapidity with which claims get paid.

[4] Sarbanes-Oxley Section 404 concerns using internal controls including audits to ensure that financial reports represent well an organization's financial status.

Medicare Administrative Contractors (Cherf, 2010), face a second set of guidelines. These auditors contract with the Centers for Medicare and Medicaid Services (CMS) to evaluate claims. They must use probabilistic sampling when making any projections from a sample onto a claims universe, including cluster, simple, or stratified random sampling (Dorfschmid, 2010). The Centers for Medicare & Medicaid Services (CMS) does not set forth a financial (or claims) accuracy rate for auditors to consider, but instead, focus on claims paid in error. According to the CMS, the Payment Error Rate is:

> An annual estimate of improper payments made under Medicaid and CHIP equal to the sum of the overpayments and underpayments in the sample, that is, the absolute value of such payments, expressed as a percentage of total payments made in the sample.(Centers for Medicare & Medicaid Services, United States Government, 2013)

In its denominator, this error rate considers total payments made, including inaccurate claims payments. One can argue that the denominator should instead use the amount of claims payments that would have been made, if in fact, claims had been paid accurately. The numerator considers incorrect payment amounts in absolute terms, a common way to consider total amount of payment errors. The resulting ratio concerns claims errors, which is not the same thing as financial accuracy.

The Healthcare Financial Management Association produced a series of Map Keys, or Key Performance Indicators (KPI's) against which Physicians Practices Management and Hospitals and Health Systems can benchmark (Healthcare Financial Management Association, 2017). Four of these concerns claims or claims adjudication, but none specifically address financial accuracy.

The question of how to measure financial accuracy—the degree to which claims payment amounts seem accurate—is an important one. From a compliance and contractual perspective, certain error rates simply may not be acceptable. Having an accurate tool for measuring an error rate helps ensure that the claims administrator meets existing contractual obligations. From a statistical perspective, financial accuracy rates define, in part, the optimal sample size. If the rate formula does not make sense, the resulting error rates will likely not be accurate, which in turn means the sample size will not be appropriate ((Agresti and Franklin, 2017). This, in turn, can result in inaccurate confidence intervals for error rates, and inaccurate estimates of precision (i.e., margins of error) (Agresti and Franklin, 2017). From a practical perspective, using an inaccurate tool wastes resources including time and money. Yet many auditors do not employ the best tool available to them to measure financial accuracy. In this article, I apply statistical principles and mathematical reasoning to make the case for adopting a particular kind of financial accuracy rate when conducting health claims audits. My perspective is that of a consultant who has consulted with a claims auditing firm (i.e., external auditors). In the interests of full disclosure, I note that a financial ratio represented here was developed and adopted by a consulting client (Claim Technologies Incorporated, or CTI); my intention is to evaluate it vis-à-vis another ratio commonly used so that others may appreciate its value.

The issues of when and how to sample also deserve attention. Some (i.e., see (Sillup, 2010)) argue for auditing all claims and not using random sampling. Such works with computer algorithms used to check for certain kinds of claims processing errors, but not all errors can be detected using computer programs. Having an auditor review every claim for payment accuracy seems needlessly expensive, since appropriate sampling and measurement procedures allow for accurate estimates that can be generalized to the claims universe.

While both the AICPA, CMS, and many states acknowledge the relevance of probabilistic sampling, the AICPA sets a very low, and arguably inappropriate standard here, supporting random sampling, but going no further.[5] The CMS also considers probability sampling like random sampling and systematic sampling when identifying claims payments errors (Centers for Medicare & Medicaid Services, United States Government, 2013). These approaches do not necessarily ensure that the most expensive errors get the most attention. Guidelines for their selection do not take into account the shape of the health insurance claims distribution.

Herein, I argue in favor of using a particular kind of probabilistic sampling—disproportionate stratified random sampling—when completing the sampling phase of health claims audits involving right-skewed claims distributions. Right-skewed claims amounts distributions, found commonly with health care data, have few large dollar claims and many smaller dollar claims. As evidence that such is true please see, for example, actuarial textbooks that have long used right-skewed examples for health insurance claims amount data (e.g., (Klugman, Panjer and Willmot, 1998), and (Comstock, 1996). Privacy agreements prevent the author from sharing health insurance claims data from their consulting work in this paper, and publicly available health insurance claims data sets from employers that can be used in this analysis are scarce, due in part to HIPAA.

Finally, regarding stratification, I explain the advantages of choosing to stratify using billed claims rather than the frequently used paid claims. The methods advocated for herein, when taken in combination, represent a departure from common practices and standards, including the vaguely phrased 'random sampling' approach required by the AICPA.

The rest of the article is organized in the following way. The next section concerns ways of calculating financial accuracy rates. The third section includes an examination of health claims screening, including the choice of non-probabilistic versus probabilistic sampling methods, simple versus stratified random sampling, and the selection of relevant stratifying variables. I present a brief discussion and conclusion in the fourth and final section.

CALCULATING FINANCIAL ACCURACY RATES

Many health claims auditors use some type of financial accuracy rate to indicate the accuracy with which claims are paid (e.g., (Benedict, 2015), (Orejudos and Paszliewicz, 2014), and (Wells, 2010)). A ratio equal to 100 percent should indicate no payment errors whatsoever, and thus serve as the benchmark to which most health plan operators should aspire. Anything less than that should reflect claims payment errors. A ratio that exceeds 100 percent would indicate better than perfect claims payments, which is logically impossible.

[5] See the definition of "Statistical Sampling" in AU-C § 530.05, Audit Sampling, AICPA, 2015, p. 492 (AICPA, 2015).

A common way to measure claims accurate payment rates is by comparing the total claims dollars paid correctly with total claims dollars processed.[6] See Figure 1 for an equation that shows this idea, and results in a financial accurate rate (FAR1). Conceptually, the approach here differs from the one taken by the Centers for Medicare & Medicaid Services (CMS) and discussed in the previous section (see (Centers for Medicare & Medicaid Services, United States Government, 2013)), because here the emphasis is on claims paid accurately rather than claims paid in error.

Figure 1: FAR1

$$\text{Financial Accuracy Rate 1 (FAR1)} = \frac{\text{Total Claims Paid Accurately in Dollars}}{\text{Total Claims Paid in Dollars}}$$

The numerator, Total Claims Paid Accurately in Dollars, can be broken down as:

$$= \text{Total Claims Paid} + \text{Claims Underpayments} - \text{Claims Overpayments}$$

Another way to state the denominator, Total Claims Paid in Dollars, is as follows:

$$= \text{Amount of Claims Paid Correctly} + \text{Amount of Claims Overpayments}$$

Underpayments will not be reflected in the denominator of this equation, because they were never paid.

The numerator in Figure 1 reflects the dollar amount of claims payments that should have been made if all claims had been paid accurately. The denominator shows the actual dollar amounts of claims paid. When the actual dollar amounts of claims paid equals the total amounts of claims paid accurately, the ratio equals 100 percent, which seems to make sense on the surface.

However, consider the situation where plan administrators have no overpayments, but some underpayments. In that case, the ratio's numerator consists of total claims paid and claims underpayments. See Figure 2.

Figure 2: FAR1 No Overpayments Case

$$\text{FAR1, no overpayments case} = \frac{\text{Total Claims Paid in Dollars} + \text{Claims Underpayments}}{\text{Total Claims Paid in Dollars}}$$

The denominator will reflect no underpayments because it consists of total claims paid in dollars, which can only include correctly paid claims or claims overpayments. If claims underpayments are more than $0, this ratio will exceed 100 percent, which simply makes no sense. How can one process more than 100 percent of claims accurately? One cannot, and so the ratio becomes less useful.

[6] See, for example, Audit of the Health Care Claims Processing for Calendar Years 2002 and 2003, City of Dallas, Texas, prepared by James R. Martin, CPA, and Tony Aguilar, CISA, published by the Office of the City Auditor, September 23, 2005, p. 3.

Consider also the practice of creating a denominator that consists of total claims paid in dollars, which typically includes some claims overpayments as well as correctly paid amounts, to use in norming against a numerator of correctly paid claims amounts. This, too, seems illogical. If the benchmark to which the plan administrator ascribes is completely accurate claims payments, why include an incorrect claims payment figure in the denominator? The only time that the denominator will reflect claims paid accurately is when there are no overpayments or underpayments. If there are underpayments but no overpayments, the denominator will be too small, making the financial accuracy rate too large vis-à-vis the real rate. If overpayments but no underpayments exist, the denominator will be too large, making the financial accuracy rate too small vis-à-vis the true rate. If both underpayments and overpayments exist, the extent to which the rate over- or underestimates the true rate depends on which is greater, over- or underpayments.

Here, a 99 percent financial accuracy rate means that the dollar amounts of total claims paid correctly accounts for 99 percent of the total claims amounts paid. It does not mean that 1 percent of claims paid amounts were not correct.

To overcome both of these difficulties, consider the following way of calculating financial accuracy rates, shown in Figure 3, and used by CTI.[7] This ratio focuses on the share of claims paid correctly, derived by subtracting a kind of claims accuracy ratio from 100%. The claims accuracy ratio can be constructed using the audit sample results, where the denominator reflects the corrected amounts of claims payments, and the numerator reflects the dollar amounts of incorrectly paid claims.

Figure 3: FAR2

$$\text{FAR2} = 100\% - \frac{\text{Dollar Amount of Incorrectly Paid Claims}}{\text{Dollar Amount of Claims if All Had Been Paid Correctly}}$$

Expanding the fraction's numerator, Dollar Amount of Incorrectly Paid Claims, yields:

$$= \text{Claim Underpayments} + \text{Claims Overpayments}$$

Where Claim Underpayments and Overpayments are expressed in positive dollar amounts (i.e., in absolute terms)

Expanding the FAR2 denominator, Correctly Paid Claims Payments Amount, gives:

$$= \text{Dollar Amount of Paid Claims} + \text{Underpayments} - \text{Overpayments}$$

Because FAR2 considers the ideal payments situation—the dollar amounts of claims that should have been paid if all payments had been made correctly—it provides a true measure of what administrators hope to accomplish: accurate claims payments. A 99 percent financial accuracy ratio should indicate that 1 percent of the dollar amounts of claims are paid inaccurately; FAR2 is logically consistent with this interpretation.

[7] The source for this formula was internal documents from CTI; I share it with CTI's permission.

The FAR2 will not exceed 100 percent. In order to rise above 100 percent, either the dollar amount of incorrectly paid claims or the dollar amounts of claims if all had been paid correctly would have to be negative. Because both items are at least 0, the FAR2 will always be less than or equal to 100 percent.

When the dollar amount of incorrectly paid claims is 0, that is to say that all claims have been paid correctly, FAR2 = 100 percent, as it should. When the dollar amounts of claims paid incorrectly is some positive amount, the FAR2 will fall below 100 percent. Both of these results make sense in practice.

Summary

In this section, I examined two approaches to calculating financial accuracy rates. The first and perhaps most common approach, FAR1, suffers from two logic problems, namely that the ratio can exceed 100 percent, indicating better than perfect claims payments, and that the ratio norms against actual claims paid, which is not the ideal against which claims payments systems should be benchmarked. The second ratio (FAR2) overcomes both of these problems, and seems just as simple to calculate as the first. It has the added advantage of producing values that are logically consistent with the usual interpretation of financial accuracy rates, namely the share of dollar amounts of claims paid accurately.

In the next section, I consider the consequences of using paid claims to stratify health claims audit samples. Both this and the next section lend insights into ways in which health care audit processes can be improved.

SCREENING CLAIMS

Health claims audits often consist of reviews for systematic errors of claims paid over a 12- or 24-month period. The number of periods is usually defined by the audit contract, since the amount of information included in an audit helps frame audit scope and thus, cost. All other factors held constant, a 24-month window works better than a 12-month frame for at least two reasons. First, there may be a lag between when eligible medical expenses are incurred and when they are reported. It is not unusual to have medical claims development three months after the end of the calendar or fiscal year (i.e., in months 13, 14 and 15). Therefore, working with 12 months' worth of data will not provide a complete view of claims for a calendar (or fiscal) year. Because medical claims can be seasonal, varying depending on the month of the year, one full year's worth of data should be reviewed at a minimum. Data drawn from a 24-month period provides more information than data drawn from a 12-month period. With the additional information comes the opportunity for more accurate data analysis. Thus, a second reason for preferring 24 over 12 months of data comes from statistical practice: more information allows for better analysis of the claims and more reliable estimates about populations, assuming that the claims population is not changing rapidly during the time period considered (see, for example, (Agresti and Franklin, 2017) for discussions on the effects of increasing sample size on the accuracy of inferences made about a population).8 Of course, using, say, a 27-month window would allow for the capture of additional year two claims development and help in creating a more accurate view of claims processing concerns for two fiscal (or calendar) years.

A counter-argument to using 24 months of data is that codes, coding rules, and contract terms can change over time. That argument does not negate the arguments that I make above, but certainly can complicate the audit.

Phase One: Electronic Screening

Most health audits occur in two phases. The first phase involves using software to conduct a highly automated review of the claims universe. The software employed varies from vendor to vendor, but often proves highly effective in detecting systematic processing errors created by programming errors (health claims administration is typically highly automated.)9 An example of a programming error is a daily hospitalization benefit maximum dollar amount entered incorrectly into claims adjudication software, which then gets applied to all relevant hospitalization claims (Wells, 2010). More generally speaking, classes of errors frequently detected through electronic data analysis include duplicate claims payments, benefit maximums errors, payments to ineligible beneficiaries, and payments for ineligible services (4 Health Care Claims Auditing Tips Help Pare Overpayments, 2008).

While the ability of software to screen for claims payment errors has improved significantly over the past few decades, it is not fool-proof. Thus, many auditors add a manual review of a sample of claims to their claims audits. Using a sample in phase-two, rather than a census of the claims universe, means that not all claims are selected for closer scrutiny. Indeed, a census of all claims would be unduly time-consuming, costly, and unnecessary if appropriate statistical methods are employed. It is possible to create reasonably accurate projections about the claims universe using sample information if good sampling procedures are employed (Agresti and Franklin, 2017).

Related to this topic is the question of whether a 100 percent accuracy rate is a reasonable goal to achieve for a plan. While raising the financial accuracy rate to levels near 100 percent makes sense, the ideal of 100 percent accuracy, exposed by Sillup and Klimberg (2010), is in many cases, impossible to achieve without a census. Achieving 100 percent accuracy means detecting all errors, which cannot be viably done through current electronic screening technology. Nearly all plans involve some errors; detecting all of these errors would require a tremendous resource outlay that would include a census of the claims universe. A census is not needed if sound probabilistic statistical sampling is employed. Results from such a sample can be used to make

[8] Staying with this line of reasoning, including even more months in the audit will provide a yet more accurate picture of claims paying performance, everything else the same. This is simply increasing the number of observations one has from the universe. See (Agresti and Franklin, 2017) for further discussion on sample effects on inference.

[9] Health care claims typically reside in a data warehouse along with other pertinent information such as eligibility files, summary plan documents, administrative services only (ASO) contracts, prescription drug pricing, American Medical Association and Food and Drug Administration (FDA) guidelines.

generalizations about the claims universe without doing an expensive and time-consuming census of the universe (see (Agresti and Franklin, 2017) for a discussion of sampling and inference).

Phase Two: Selecting Claims for Closer Scrutiny

A second phase includes manual reviews of a smaller number of claims. The phase two sample should be drawn from the claims universe, and not just a subset of claims that were flagged in phase one. Not all claims with errors will be flagged in phase one, because not all errors lend themselves well to detection using software screens. Coordination of benefits problems often go undetected in phase one, for example. Also, manual adjudication, frequently used for large, complex claims (Wells, 2010), often creates errors, some of which may go undetected using screens reserved for standardized claims processing.10 Errors on large dollar claims usually prove to be the most economically important errors observed in an audit, so particular attention should be paid to detecting them using manual reviews. If reviews are confined to only those claims that were flagged during the automated phase one, some errors will get overlooked.

For those who correctly choose to sample, the process used to select claims to be included in the phase two audit seems to vary. Many employ non-probabilistic sampling, using something other than chance to identify claims that may require further scrutiny (Elder, et al., 2013). Others use probabilistic sampling, allowing chance to determine which claims get selected for manual review during phase two. I further describe both types of sampling below.

Non-probabilistic Sampling. Chance is not used to determine which claims receive further review with non-probabilistic sampling.11 This kind of sampling presents at least two problems. First, Elder et. al., (2103) note that while non-probabilistic audit sampling is common, auditors are "prone to decision biases" when evaluating their results (Elder, et al., 2013). Second, when compared with estimates from probabilistic samples, estimates from non-probabilistic samples are more likely to suffer from bias and thus are less accurate estimates of claims universe measures (see, for example, (Agresti and Franklin, 2017) for a more general discussion of sampling and statistical inference).

While auditors can calculate measures using sample data and make statements about them, like the sample's financial accuracy rate was 97 percent, those statements simply describe the sample. Using nonrandom sample results in making generalizations from the sample to the claims universe should be avoided when using non-probabilistic sampling. In practice, this means that an auditor should not say that the health care plan's financial accuracy rate was 97 percent (or 97 percent plus and minus some margin of error) if a nonrandom sample was used, because the

---

[10] The more complex the claim, the greater the chance that an error is made in settling it, every other factor held constant.

[11] If, say, your phase two sample consists of all claims flagged as claims involving errors during a phase one electronic screen, you are employing nonrandom sampling. As an aside, these claims may present no other errors other than those determined through a phase one screen. Thus, further scrutiny of these claims may be a complete waste of time. Meanwhile, claims paid in error for reasons other than software coding issues will not be at all addressed in the phase two audit.

sample most likely does not represent well the population so statistics calculated from it will not likely represent well the claims universe under consideration (i.e., the sample financial accuracy rate will likely not estimate accurately the true financial accuracy rate for the claims universe).

Simple Random Sampling. Other auditors use simple random sampling to select claims for phase two evaluations. Simple random sampling relies on chance to determine which claims get scrutinized in phase two, a much better approach than non-probabilistic sampling for creating a sample that represents well the claims population of interest. In theory, measures from a simple random sample can be used to describe a claims universe (i.e., the health care plan's financial accuracy rate was 97 percent, plus and minus 1 percent). Yet these numbers may not be very accurate because simple random samples often fail to capture the most expensive claims payments errors. Because most medical claims are small, typically less than $500, most claims selected through random sampling will also typically be small. It follows that a payment error on such claims will also usually be small (i.e., a $250 lab bill that accidentally omitted $15 co-pay results in a $15 payment error). Because large claims occur much less frequently, few will likely be included in a random sample drawn from the medical claims universe. In instances where large claims account for a very small share of the total number of claims, it is possible that none will be included. Note that a payment error on a large claim is less likely to be small in magnitude, even if it accounts for a small portion of the total claim expense. While a 10 percent error for a $500 claim equals $50, the same percentage error for a $500,000 claim equals $50,000. A 10 percent error is not particularly large in percentage terms, but a $50,000 claims payment error is large in dollar terms. The lesson here is that careful scrutiny of some of the largest claims should help in identifying the largest payment errors. A properly selected simple random sample may not produce the "right sample" to identify errors in processing the largest claims. A better approach is to use stratified random sampling.

Stratified Random Sampling. Stratified random sampling is appropriate when layers or strata exist within a population (or universe, as populations are commonly called in health care claims auditing) and each layer must be represented within a random sample. In the case of claims payments, those layers are typically expressed in financial terms, either as claims dollars paid, or billed.12 Using some measure of claims payments in stratifying ensures that claims will be drawn from a variety of claims amounts, helping to ensure that large claims, which usually occur far less frequently than small ones, have representation in the sample (i.e., when a claims distribution is right-skewed, a common shape for a health claims distribution, as evidenced by its use in examples in actuarial textbooks like (Comstock, 1996) and (Klugman, Panjer and Willmot, 1998)). For now, I will assume that stratifying ought to be done using billed amounts rather than actual paid claims. A discussion of stratification approaches—proportionate versus disproportionate—will help the reader understand why such is the case. That discussion comes next, followed by a

[12] In a review of allowable Medicaid patient days, one auditor attempted to stratify using the length of patient stay for three of the four strata employed. The Provider Review Reimbursement Board (PRRB) held that because a patient's length of stay was not related to a patient's Medicaid eligibility, it should not be used to stratify associated audit samples (Kerry, 2011).

more complete set of arguments explaining why billed amounts should be used to stratify in lieu of actual paid amounts.

Stratified random sampling (SRS) may be proportionate or disproportionate. With proportionate stratified random sampling, the number of all claims accounted for by a layer divided by the total number of claims in a population or universe defines that layer's share of the sample. In health care claims audits, billed dollar amounts might be used to construct layers. Suppose that there are 1,000 claims in a population, and 700 of these have total charges of $500 or less; the first layer accounts for 70 percent of the population claims, and thus, 70 percent of the sample claims. It follows that if the total desired sample size is 100, 70 of the claims sampled should come from the first layer (i.e., billed amounts less than or equal to $500).

With disproportionate stratified random sampling, the layer's share of claims in the sample (e.g., approximately 30 percent) is something other than the share of claims accounted for by that layer in the population (e.g., 70 percent). Disproportionate stratified random sampling allows the auditor to over-sample one layer that may be more important financially (e.g., billed amounts exceeding $5,000), and under-sample another layer that may have less of an impact on the bottom line (e.g., billed amounts less than or equal to $500).

Proportionate SRS works better than disproportionate SRS when the variability in the stratifying variable is approximately the same throughout the layers, but not when it differs significantly (Thompson, 2002). The variability of billed claims typically differs across layers, with less variation observed among lower billed amounts than among larger charged amounts. Estimates of errors expressed as billed dollars will therefore be more precise with disproportionate SRS than with proportionate SRS.

There is another reason why disproportionate stratified random sampling should be used in auditing health claims. Because claims processing errors in the high dollar strata will often have a much greater effect on the bottom line than errors for smaller claims, an auditor should over-sample (i.e., select a larger number of claims from this layer than would be warranted if one were using proportionate SRS) from this layer. Errors made in processing high dollar claims usually represent the greatest potential sources of cost savings to audit clients. By over-sampling from these claims, an auditor likely improves significantly their ability to detect the costliest errors a client may incur and correct those.

Note that over-sampling increases the chances of discovering more expensive errors. Thus, the practice can adversely affect the estimated financial accuracy ratio, causing it to be lower than it might be with simple or proportionate random sampling. If the goal is to have the estimated highest financial accuracy ratio possible, an auditor should not over-sample from the costliest claims layer. In fact, an auditor should probably not use stratified random sampling at all, as simple random sampling, which does not ensure that any large claims will appear in the sample, should produce a more favorable estimated financial accuracy ratio, all other factors held constant. However, if the goal is to identify process improvements that can help ensure accurate claims payments, and thus help ensure that an organization is not spending too much (or too little) on claims, then one can over-sample in the highest billed claims and under-sample in the lowest.

Having examined the usefulness of simple random sampling, and proportionate and disproportionate stratified random sampling in conducting the second phases of claims audits, and making a case for using disproportionate random sampling, I now return to the selection of appropriate stratifying variables.

SELECTING APPROPRIATE STRATIFYING VARIABLES

Many auditors stratify audit samples using paid claims amounts. Yet a better way to discover claims payment errors, especially among large claims, is to stratify using billed charges. To understand why, it is helpful to first consider the differences between billed and paid claims.

A claim is created when a healthcare provider submits a request for payment for services rendered to a payer, creating a billed claim. The bill may be paid in full or in part, or completely denied which results in a $0 paid claim. The denial may be legitimate, an accidental error, or, sadly enough, an intentional strategy to avoid paying costly claims.

If the billed amount equals the paid amount, then no discrepancy exists between billed and paid claims amounts, and sample stratification by either billed or paid claims would yield the same results. However, if a discrepancy exists, stratification using the two variables will not necessarily produce the same results. In particular, legitimate claims that are completely denied will have $0 paid amounts, and thus end up in the lowest and, when disproportionate stratified random sampling is used, under-sampled dollar stratum, where they are less likely to be included in any audit sample. If these legitimate claims have large billed amounts (e.g., bills of at least $5,000), including them in the largest and oversampled strata increases the chances that they will be included in the phase two audit sample (Gagne, 2012). If the auditor's aims include encouraging accurate claims payments, they should use billed rather than paid claims to stratify (Gagne, 2012).

Interestingly enough, many and perhaps most auditors still use actual dollars paid in claims to stratify (see, for example, (Cohen, 2007), (Auditors, 2007), or (Washington State HealthCare Authority, 2017)) or to select their sample (BMI Audit Services n.d.).) This practice is problematic for the reasons described above.

Summary

Determining the best way to screen claims when conducting health claims audits requires careful planning. Decisions about the number of months of claims that ought to be reviewed, how many phases should be employed in screening claims, and how claims ought to be selected for further review may all influence heavily an audit's outcome. In this section, I make the case for using a minimum of 24 months worth of claims, rather than a shorter time frame. More months of claims provide more information, and thus, more credible results (for the statistical reasoning behind this statement, see, for example, (Agresti and Franklin, 2017)). I also discuss a two-phase approach for screening claims, with the first phase relying on electronic screens which can be helpful in detecting certain kinds of errors, but not all errors. The second phase should include drawing a stratified random sample from the entire claims universe using billed claims amounts to ensure that the largest, most expensive claims get the kind of scrutiny that they deserve. The approaches that I recommend in the second phase depart from those that seem to be commonly employed in audits (see, for example, (Cohen, 2007), (Auditors, 2007), or (Washington State HealthCare Authority, 2017). They are consistent with existing AICPA guidelines, but those guidelines are written very broadly and thus will also include other, less satisfactory approaches to health claims audits.

CONCLUSION

In this article, I make the case for a different approach to calculating financial accuracy rates than what is commonly used. I do this by comparing a common way of calculating a financial accuracy rate, which compares total claims paid accurately with total claims paid, with a ratio that considers the amount of incorrectly paid claims vis-à-vis the amounts of claims that would have been paid if all claims had been paid accurately. The later ratio proves more logically consistent with the notion of financial accuracy, and yields mathematically appropriate values bounded always by 0 percent at the low end and 100 percent at the high end.

I also make a case for using stratified random sampling rather than simple random sampling when conducting audits of health claims. Stratified random sampling helps to ensure that low frequency, high expense claims get included in the audit sample. Errors on these claims can be particularly costly. Therefore, these claims deserve closer scrutiny than, say, $20 claims.

While many do use stratified sampling, they often do not stratify using billed rather than paid amounts. Herein, I argue that claims should be stratified using billed rather than paid amounts. The literature indicates that most auditors use paid amounts, which means that if a large claim is denied by a health insurer, the paid amount is $0, possibly giving the auditor the false impression that the claim is not important economically. This same large claim would be recorded as an expensive claim using billed amounts, and thus, seem more economically important, and thus deserving of more scrutiny during an audit.

REFERENCES

Agresti, A., & Franklin, C. (2017). *Statistics: The Art and Science of Learning from Data* (4th ed.). Boston, MA, US: Pearson.

AICPA.(2015). *https://www.aicpa.org/Research/Standards/AuditAttest/DownloadableDocuments/AU-C-00530.pdf.* Retrieved September 20, 2016, from Audit Sampling.

Auditors, A. (2007). *Texas Health and Human Services Commission and Clarendon National Insurance Company.* Compliance Examination Report.

Benedict, M. (2015, November 12). *How to Gather Good Data for Healthcare Claims Accuracy Testing*. Retrieved October 5, 2017, from Health IT Analytics: https://healthitanalytics.com/news/how-to-gather-good-data-for-healthcare-claims-accuracy-testing

BMI Audit Services. (n.d.). *Health Care Claims Audit Sampling Methodologies Random vs. Judgmental .* Retrieved October 8, 2017, from BMI Audit: https://bmiaudit.com/pdf/BMI-Audit-Services-Random-Judgemental-Health-Care-Claim-Audits.pdf

Bohnsack, J., & Hawig, S. (2012, September). Choosing the right strategy for point-of-service collections: technologies that accurately predict a patient's financial responsibility, combined with effective staff training and benchmarks for collections performance, can enhance revenue and satisfaction. *Healthcare Financial Management, 66*(9), 122-128.

Centers for Medicare & Medicaid Services, United States Government. (2013, October 15). *Payment Error Rate Measurement Manual.* Retrieved October 5, 2017, from https://www.cms.gov/Research-Statistics-Data-and-Systems/Monitoring-Programs/Medicare-FFS-Compliance-Programs/PERM/Downloads/CMSPERMmanual.pdf

Cherf, K. M. (2010). The Facts on RACs. *Health Law Litigation*, 1-10.

Claim Technologies Inc. (2007). Des Moines, Iowa: Claim Technologies Inc.

Claim Technologies Incorporated. (2008). A CTI White Paper. *Audit and Control of Health Plan Claims Administration*. Des Moines, Iowa: Claim Technologies Incorporated.

Cohen, A. a. (2007). A representative sampling plan for auditing health insurance claims. *Lecture Notes - Monograph Series, Vol. 54, Complex Data Sets and Inverse Problems: Tomography, Networks and Beyond*, pp. 121-131.

Comstock, S. J. (1996). Estimating Claim Costs for Traditional Health Benefits. In W. F. Bluhm, A. D. Ford, J. E. Lusk, & P. L. Perkins (Eds.), *Group Insurance* (Second ed., p. 394). Winsted, CT, US: ACTEX Publications.

Cusick, D., & Anderson, B. (2014, June 4). *Do Not Overlook the Value of Claims Auditing*. Retrieved October 5, 2017, from Employee Benefit News: https://www.benefitnews.com/news/dont-overlook-the-value-of-claims-auditing

Dorfschmid, C. M. (2010). Why RAT-STATS and Sampling are Hot. *Journal of Health Care Compliance*, 41-46.

Elder, R. J., Akresh, A. D., Glover, S. M., Higgs, J. L., & Liljegren, J. (2013). Audit Sampling Research: A Synthesis and Implications for Future Research. *Auditing: A Journal of Practice & Theory, 32*, 99-129.

Four Health Care Claims Auditing Tips Help Pare Overpayments. (2008, May). *Managing Benefits Plans*, pp. 1, 13-15.

Gagne, P. a. (2012, July). Discussion regarding stratification by paid claims amounts. (L. Gardner, Interviewer)

Healthcare Financial Management Association. (2017). *mapkeys*. (H. F. Association, Producer) Retrieved October 2015, 2017, from hfmap: https://www.hfma.org/MAP/MapKeys/

Hurlock, W. (2010, November). The need for health care paid claims and pharmacy benefit audits: what you don't know may hurt you. *Employee Benefit Plan Review, 65*(5), 5-6.

Kalish, B. (2013, April 1). *Dependent eligibility audits mean big cost savings, big disruption*. Retrieved October 5, 2017, from Employee Benefits News: http://bi.galegroup.com/global/article/GALE%7CA324368359/264505158b78a34c37b911a2b901bbcd?u=drakeu_main

Kerry, C. (2011, September). PRRB Decision Places Boundaries on Contractor Sampling Techniques. *Reimbursement Auditor*, pp. 7-10.

Klugman, S. A., Panjer, H. H., & Willmot, G. E. (1998). *Loss Models: From Data to Decisions.* New York, NY, US: Wiley .

Lubbers, M. A. (2010). *Healthcare Data Audit.* Cincinnati, Ohio: City of Cincinnati.

Martin, J. R. (2005). *Audit of the Health Care Claims Processing for Calendar Years 2002 and 2003.* Dallas, Texas: Office of the City Auditor of Dallas, Texas.

Mcguffie, D. L. (2011, April). Taming the beast: reining in health-care fraud: through audits of areas such as dependent eligibility and claims, auditors can help stem losses that wreak havoc on the organization's bottom line. *Internal Auditor, 68*(2), 28-33.

Orejudos, F., & Paszliewicz, D. (2014, September 4). Converting Claim Audit Findings into Cost Savings. *Self Funding Magazine*(43).

Ramunni, K. (2016, February 16). *Leng orders audit of Hamden's medical expenses*. Retrieved October 5, 2017, from New Haven Register: http://go.galegroup.com/ps/i.do?&id=GALE|A443343529&v=2.1&u=drakeu_main&it=r&p=ITOF&sw=w&authCount=1

Sillup, G. P. (2010). *Health Plan Auditing: 100-Percent-of-Claims vs. Random-Sample Audits Look at What You're Missing.* White Paper, Philadelphia, PA.

State News Service. (2016, December 28). *Home Health Care Group Sues CMS Over Pre-Claim Audit Pilot Program.* Retrieved October 5, 2017, from Academic One File: http://go.galegroup.com/ps/i.do?&id=GALE|A475218333&v=2.1&u=drakeu_main&it=r&p=AONE&sw=w&authCount=1

Stegman, M. S. (2008). Recovery Audit Contractor Review Update: CMS Plans to Gradually Expand Nationwide by 2010. *Journal of Health Care Compliance*, 49-52.

Thompson, S. K. (2002). *Sampling* (2nd ed.). New York, New York, US: Wiley Series in Probability and Statistics.

Washington State HealthCare Authority. (2017, June 1). *Sampling and Extrapolation Process For Section of Program Integrity (PI) .* Retrieved October 8, 2017, from https://www.hca.wa.gov/assets/program/sampling-extrapolation-process.pdf

Wells, S. J. (2010, July). Checking Up On Medical Claims. *HR Magazine*, p. 38.