1996

Third Party Administrator (TPA) Service Pricing and Incentive Contracts

Hou-Wen Jeng
Nationwide Insurance, jengh@nationwide.com

Follow this and additional works at: http://digitalcommons.unl.edu/joap

Part of the Accounting Commons, Business Administration, Management, and Operations Commons, Corporate Finance Commons, Finance and Financial Management Commons, Insurance Commons, and the Management Sciences and Quantitative Methods Commons

http://digitalcommons.unl.edu/joap/112

This Article is brought to you for free and open access by the Finance Department at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Journal of Actuarial Practice 1993-2006 by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
Third Party Administrator (TPA) Service Pricing and Incentive Contracts

Hou-Wen Jeng*

Abstract†

This paper addresses a few of the most important pricing issues faced by a third party administrator (TPA) whose main responsibility is claims handling for self-insured employers and self-insured groups. Such pricing issues include the development of service fees using claim closure information, the selection of service durations, and the design of incentive (either activity-based or financially-based) service contracts. Formulas for pricing new and open claims are provided.

Key words and phrases: self-insurance, service length, new claims, open claims

1 Introduction

Self-insurance programs are designed to capture the potential cash flow benefits arising from loss reserves and expense savings. To achieve these goals, self-insured employers and self-insured groups need to carefully select a professional service provider, also known as a third party administrator (TPA). TPAs have substantial experience in claims handling, and they usually have access to other supporting services.

*Hou-Wen Jeng, A.C.A.S., A.Re., is an actuary with Nationwide Insurance. He holds a B.A. in economics from National Taiwan University and an M.A. in economics from the University of Rochester.

Mr. Jeng’s address is: Nationwide Insurance Co., One Nationwide Plaza, Columbus OH 43216, USA. Internet address: jengh@nationwide.com

†This paper is based on a previous paper entitled “TPA Service Pricing and Incentive Contracts” that appeared in Alternative Markets/Self Insurance—1996 Discussion Paper Program, a nonrefereed publication of the Casualty Actuarial Society. The author thanks Rich Cundy, Rudy Palenik, Virginia Price, the editor and the anonymous referees for their comments. All remaining errors are his responsibility.
such as actuarial, loss control, managed care, and return-to-work programs. Thus, a TPA generally is regarded as the centerpiece of many self-insurance programs.

From a service standpoint, the most significant difference between a TPA and a claims department of an insurance company is that a TPA provides claims services with a variety of service lengths, ranging from twelve months to the life of the claim. The primary product lines for self-insurance are workers' compensation and general liability which are also considered long-tail lines in insurance. Long-tail claims not only complicate the pricing for TPAs, but greatly affect the TPA fee options and service lengths available on the market. Given the long-tail nature of the product lines and the variety of the service lengths, TPAs in general have difficulties in forecasting the costs and pricing their products.

Techniques used in insurance ratemaking and reserving may shed some light on how TPA service pricing should be performed. The aggregate approach used in insurance regarding unallocated loss adjustment expenses (ULAE), however, is not appropriate for pricing TPA products. A more detailed approach using service time and closing ratio by claim age works well in predicting claim handling costs for various service lengths. Here we emphasize the significance of using claim age in the service fee development. Service level is assumed to be a function of claim age. The distribution of claim ages is related to claims closure distributions. This paper illustrates how information can be combined in the development process.

The last pricing issue discussed is the design of incentive contracts. This has become increasingly important for TPA pricing, especially in financial incentive contracts, due to surging market demand. Two major types of performance measurements for incentive contracts are discussed, and a recommendation is made considering factors that impact the financial results of a self insurance program.

When discussing TPA pricing procedures and incentive contracts, the paper focuses on workers' compensation. The formulas, procedures, and examples can be generalized to include other lines such as general liability and auto liability.

2 Fee Options and Service Length

TPA service pricing is not examined as closely by state regulatory agencies as is insurance pricing. State agencies may assume that, like reinsurance pricing, both parties are large and knowledgeable regarding
the trade in which they are engaged. As a result, the pricing of TPA service contracts is extremely competitive and TPAs usually customize their products to fit the needs of clients.

A TPA typically is expected to provide several service fee options, including per claim, dedicated office/unit, percent of incurred, and percent of paid. There may be one or more choices of service length for each of the fee options, ranging from 12 months to the life of the claim. Table 1 lists the major TPA service fee options and the service lengths available for the corresponding fee option.

Table 1
Major TPA Service Fee Options

<table>
<thead>
<tr>
<th>Fee Options</th>
<th>Service Length Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Per Claim</td>
<td>12 months</td>
</tr>
<tr>
<td></td>
<td>24 months</td>
</tr>
<tr>
<td></td>
<td>Life of partnership</td>
</tr>
<tr>
<td></td>
<td>Life of claim</td>
</tr>
<tr>
<td>Dedicated Office/Unit</td>
<td>Same as contract period</td>
</tr>
<tr>
<td>% of Incurred/Paid Loss</td>
<td>Usually life of claim</td>
</tr>
<tr>
<td>% of Premium</td>
<td>Usually life of claim</td>
</tr>
<tr>
<td>% of Employees</td>
<td>Usually life of claim</td>
</tr>
</tbody>
</table>

Before elaborating on the different TPA service fee options, it is helpful to discuss several pricing-related risks: claim frequency, claim severity, loss development, and premium adequacy. Per claim involves two risks: claim severity and loss development. TPAs can mitigate the risks by charging different fees for different types of claims if such a classification is feasible. For example, in workers' compensation, the service fees for medical-only and indemnity claims are significantly different.

If the service fee is based on a percentage of premium, then the TPA faces three risks: frequency, severity, and premium adequacy risks. In other words, the TPA has to absorb the servicing cost of any unexpected increases in claim frequency and severity, and needs to make sure that

---

1. Very rarely do TPAs charge by hours of time, especially for long tail lines because it is difficult to pinpoint the exact time spent for each claim.
2. Here we only consider the pace of the loss development.
the calculated premium\textsuperscript{3} is adequate. On the other hand, if the service fee is based on a percentage of paid/incurred losses,\textsuperscript{4} the TPA needs to be concerned with the pace of the loss development (although the risks of frequency and severity may be smaller than the premium-based pricing). In practice, many TPAs wish to assume as little insurance-related risks as possible. The fear is that they may be considered as insurers and be regulated as such. In addition, they may not have the resources and insurance expertise to underwrite insurance risks.

2.1 The Per Claim Fee Option

2.1.1 Basics

Because it is flexible in service length, the per claim fee option has been the most popular choice among self-insureds, where service fees are based on the number of claims received by the TPA in the contract period. Under per claim, a self-insured client can choose from various service lengths for the claims to be serviced, such as 12 month, 24 month, life of partnership, and life of claim. This diversity in service length contrasts with traditional insurance where insurers always service claims to conclusion.

The fee for the 12 month claims service provides claims handling on new claims reported during the contract period and claims open at the beginning of the contract period for a period of 12 consecutive months. Similarly, the 24 month claims service provides claims handling for 24 consecutive months. Consider an example where the contract period is from 1/1/95 to 12/31/95 and 24 months is the selected service length. A claim reported on 3/1/95 will be serviced continuously until 2/28/97, 14 months after the end of the contract period. Similarly, a claim reported on 7/20/95 will be serviced continuously until 7/19/97. The total fee calculation using the data in Table 2 is simple:

\[
\text{Total Fee Charges on 12/31/95} = 250 \times 200 + 550 \times 300 = 215,000.
\]

For a new customer, the charges for the open claims assumed at the inception of the contract can be easily determined and billed. New claims (i.e., claims that have never been serviced by a claims administrator) are billed only when they are reported to the TPA. As a result,

\textsuperscript{3}Most self-insureds report their payrolls and incurred losses to the state. They do not calculate their premium, and their exposures usually are not properly classified as required in insured cases.

\textsuperscript{4}See Section 2.3 for a more detailed discussion.
Table 2
Data for a New Customer

<table>
<thead>
<tr>
<th>Contract Period</th>
<th>CY 1995</th>
</tr>
</thead>
<tbody>
<tr>
<td>Service Length</td>
<td>24 months</td>
</tr>
<tr>
<td>Per Open Claim Charge</td>
<td>$250</td>
</tr>
<tr>
<td>Per New Claim Charge</td>
<td>$550</td>
</tr>
<tr>
<td>Number of Open Claims Known as of 1/1/95</td>
<td>200</td>
</tr>
<tr>
<td>Number of New Claims During Contract Period</td>
<td>300</td>
</tr>
</tbody>
</table>

CY Z = Calendar Year Z = 1/1/Z - 12/31/Z.

the total service charges under per claim cannot be determined until the end of the contract period. The billing process can become complicated when a customer chooses different service lengths from contract to contract. Consider the following per claim contracts (given in Table 3) for a new customer starting in 1995:

Table 3
Sample Contracts

<table>
<thead>
<tr>
<th>Contract Period</th>
<th>Contract 1</th>
<th>Contract 2</th>
<th>Contract 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Service Length</td>
<td>24 months</td>
<td>12 months</td>
<td>12 months</td>
</tr>
</tbody>
</table>

CY Z = Calendar Year Z = 1/1/Z - 12/31/Z

Contract 1 and contract 2 have different service lengths. New claims reported in 1996 and 1995 will be billed as open claims in 1997 if they remain open on their first and second anniversary dates, respectively. In addition, all the open claims assumed at the inception of contract 1 will be billed again if they are open on 1/1/97. Because the service length for contract 3 is 12 months, they will be available for billing again on 1/1/98 if they are not closed by then.

To make the per claim billing process more complicated, a client can choose different service lengths for new and open claims by location and contract. To ensure receiving proper credits, TPA billing must be claim-specific and should track the status of individual claims including service length and claim anniversary date. In practice, if the current contract is not renewed, it is common for TPAs to cease servicing all
claims at the end the current contract period unless the service length is life of claim.\textsuperscript{5}

Life of claim services provide claims handling until settlement at a fixed cost for new and open claims reported to the TPA during the contract period. Life of partnership services are the same as life of claim services except the TPA will stop servicing all existing open claims if the contractual relationship between the TPA and the self-insured ceases. Due to competitive pressure, some TPAs may sell life of partnership service under the guise of life of claim service with a lower price to gain customers. Self-insureds should study the language of their service contracts regarding service length to avoid the consequences of this confusion.

\subsection*{2.1.2 Issues}

Self-insureds can reduce claims-servicing cost by choosing a service length that best fits their self-insurance program. For example, if a self-insured finds that most of its claims can be closed within two years, a 24 month service plan may be the best choice. A tail claim service\textsuperscript{6} can be purchased to handle any remaining open claims after two years of service. On the other hand, from a TPA's perspective, the longer the service length, the more uncertainty in service pricing and revenue accrual. Thus, to avoid adverse selection, a TPA needs to determine appropriate pricing relativities between different service lengths, to investigate the closure patterns of prospective clients, and to impose risk charges for longer service lengths.

Similar to unearned premium reserves in insurance, portions of the TPA revenue from a service contract need to be deferred when the service length runs across two or more calendar years. The straight-line method used in calculating unearned premium reserves cannot be applied to the calculation of TPA service fee deferrals because of the uneven service levels at the various development ages of a long-tail claim.

\textsuperscript{5}Surprisingly, there is rarely a fee adjustment for services that have not been performed. The reason may be that most self-insureds do so voluntarily. In workers' compensation, using two TPAs (one for existing open claims and one for new claims from the same work site) may cause significant confusion. On the other hand, it would not be in the interest of the self-insured to have the same TPA handle their existing open claims due to a lack of financial incentives on the part of the TPA. It is a windfall to the TPA as fewer services need to be performed. I believe the financial effect is not significant, however, because the majority of the cancellations are 12 month service contracts. Nevertheless, one can easily factor in this effect in the pricing formula using a historical cancellation rate.

\textsuperscript{6}A tail claim service handles all remaining open claims to conclusion.
As a general rule, the older the claims, the less time they need for service. But a more relevant question to ask is how much of the service fees should be deferred? To answer the question, one must know the claim closure distribution and the average amount of time examiners spend on the claim. The pricing procedure discussed below uses this information in determining service charges for per claim. The deferral percentages can be calculated from this procedure.

For contracts with long service length, casualty actuaries can provide valuable services in the areas of TPA pricing and revenue deferral. Most self-insureds, however, are just as uncomfortable as TPAs in entering a contract with a long service length. In practice, 12 month handling is the predominant choice by self-insureds for their TPA service contracts. This phenomenon can be attributed to the following three reasons:

- Because most self-insureds are generally cost conscious, the selection of a shorter duration service plan can help their cash flow.

- Shorter service durations make it easier for a self-insured to move its program to another TPA if it is not satisfied with the current TPA's services.

- When the service contract for future claims between a TPA and a self-insured is not renewed, it would not be in the interest of the self-insured to have the same TPA handle its existing open claims due to a lack of financial incentives on the part of the TPA. In the case of life of claim handling, the self-insured and the TPA need to be in close contact regarding claims handling for many years after the termination of the service contract.

From a TPA's point of view, a contract with a short service length does have its down side. More components such as the handling of the remaining open claims from prior contracts must be negotiated at the contract renewal, and renewal negotiations occur more frequently. As a result, TPA's overhead expenses may be significantly increased. If the majority of the TPA contracts have short service length, it would be difficult for a TPA to project its future claim volumes and revenues.

2.2 Dedicated Office/Unit

Dedicated office/unit is an option where a TPA establishes a claims office or a claims unit to exclusively handle claims for the client. The set-up cost and the subsequent administrative costs, as well as the TPA's overhead and profit, are paid by the self-insured.\(^7\) Under this

\(^7\)This resembles mark-up or cost-plus pricing.
option, the service length for all claims, regardless of age, is the same as the contract period. If the contract is not renewed, the TPA will stop servicing all claims at the end of the current contract period. This option poses the least pricing risk to a TPA as it has none of the frequency, severity, loss development, and premium adequacy risks, and expenses are billed as soon as incurred. This option is usually more expensive, however, and is only recommended for larger self-insureds.

To self-insureds, the major advantage of such an arrangement is that claims examiners are familiar with the self-insured and thus are able to satisfy the client's special needs in claims handling. In addition, the location of the dedicated office can be selected strategically so that most of the current and potential claimants can be in the vicinity of the claim office. This is especially beneficial to clients such as municipalities and school districts that are geographically concentrated.

An insurance company theoretically can minimize its total payout by allocating its resources between losses and adjustment expenses. Doing so recognizes that spending more on loss adjustment may reduce loss payments and potentially can result in a lower overall cost because of the better claims management.

By being self-insured and choosing the dedicated office/unit option, a customer controls its resource allocations and is able to dictate the degree of care and the amount of time examiners spend on each case. One can demand more claim examiners to service a fixed number of claims (i.e., a lower caseload per examiner) and thus provide better service to claimants. Others may opt for a higher caseload per examiner to save adjustment expenses. Thus, under dedicated office/unit, the role of the TPA is reduced to providing the staff, computer systems, and other related technical services while the client makes the more important financial decisions and determines the extent of the claims services.

### 2.3 The Percentage Approach

Based on a predetermined percentage of the base figure (e.g., incurred loss) this fee option includes three major varieties: percent of incurred loss, percent of paid loss, and percent of premium. The service length is usually the life of the claim, as it would be difficult to determine the service fee by claim age. In the case of percent of incurred loss:

---

8Compared to other pricing options, this option involves little, if any, insurance-related risk. Thus, the pricing risk refers to general business risks such as employees, rents, equipment leases, and so on. For example, when the contract cancels, the TPA still needs to pay the rents and salaries for a certain period of time.
loss, one must determine the pace of the incurred loss development in order to price a 12 month service contract. Such information is usually not available from the self-insured. Percent of premium is used less frequently than the other two, perhaps because this option requires more information and insurance expertise for underwriting.

Both percent of paid loss and percent of incurred loss are highly individualized pricing approaches, where service charges for any claims are directly related to the cost of claims. A TPA must monitor the paid or incurred amount to determine if additional billings are necessary. Consider a claim whose ultimate cost is initially estimated at $30,000. Later it is found that a medical treatment is needed for an additional $20,000. Assuming the TPA fee is set at 7 percent of incurred loss, the fee charge for this claim will increase from $2,100 to $3,500 due to the medical treatment.

From the outset, it appears that both methods are equitable ways to determine compensations for TPA services if the percentage is selected appropriately. A closer look, however, reveals that there are serious drawbacks inherent in the methods. First, the perception of a TPA as an independent third party in claims handling could be lost because TPA service fees are linked to the settlement amount. TPAs may have little incentive to control claim costs. Second, it is also difficult for TPAs to manage the billing because incurred and paid amounts for individual claims change constantly. Third, although for any claims the paid amount eventually equals the incurred amount, the timing of the claim payments under percent of paid dictates how quickly the TPA can bill its clients. For example, most of the claims in litigation are not paid until the legal issues are resolved. At the same time, most of the handling service work on those claims has already been done. Thus, depending upon the underlying frequency and severity distributions, the use of percent of paid may result in significant risk-taking on the part of the TPA in terms of potential cash flow problems.

3 Development of TPA Service Fees

3.1 Insurance Ratemaking and Reserving Considerations

In insurance ratemaking and reserving, unallocated loss adjustment expenses (ULAE) are estimated on an aggregate basis. For example, the provision for ULAE in insurance rates generally is assumed to be a certain percentage of the premium using industry experience. The reserves for ULAE usually are estimated using the ratio of the historical
ULAE to loss and allocated among individual accident years. In the annual statement of insurance companies, the ULAE reserve calculation is based on the assumption that 50 percent of the ULAE is paid when the claim is reported and the other 50 percent is paid when the claim is settled.

There have been few changes in the ways that ULAE is built into rates and how ULAE reserves are calculated. There appears to be no need for insurance companies to establish a higher level of accuracy in the estimation of ULAE. After all, the provision for ULAE accounts for, on average, only 6 percent of the rate, and the variations in loss generally overshadow those in ULAE.

On the other hand, because a TPA's major business is claims handling, the ability to break the cost down by claim type and service length is important to the pricing of TPA services. The aggregate approach and the ad hoc rules used in insurance ratemaking and reserving are inadequate for TPA service pricing.

The following section describes an approach using service time and closing ratio by claim age to predict per claim handling costs for various service lengths. The major assumption is that service level is a function of claim age. The objectives are to keep the model simple and to explain most variation in service level. One may argue that service level also depends on other factors, such as the seriousness of the claim. If such factors are also correlated with claim age, however, the assumption has implicitly considered them.

3.2 Per Claim Pricing

3.2.1 New Claims

We will now explore how claims closure and service level information can be used to develop per claim service fees. Service level (i.e., examiner time) is assumed to be a function of claim age.

Let \( t (t = 0, 1, 2, \ldots) \) be claim age measured in months and \( F(t) \) be the cumulative percentage of closed claim at the start of the \( t \)-th month, with \( F(0) = F(1) = 0 \). Thus, \( (1 - F(t)) \) can be viewed is the probability that a claim will be open at the beginning of the \( t \)-th month since it was

---

\(^9\)Claim age or time are not the only factors that should be considered. The emphasis is whether the explanatory factor can be objectively measured and if the related data are readily available for estimation. If the TPA has a consistent claims practice, claim age seems to be a natural choice.

\(^{10}\)The analysis that follows can easily be adjusted to deal with claim age measured in other time units such as quarters or years.
reported to the TPA. Let $g(t)$ be the average service time (measured in hours) spent on an open claim at age $t$ (i.e., number of hours examiners spend on a case at age $t$ months.)\textsuperscript{11}

The shape of $g(t)$ may take many forms depending on the line of business and the type of claims. Two types of service time curves are usually observed in the case of workers' compensation indemnity claims: a downward sloping curve and a humped curve with its peak within the first six months (see Figure 1). Both curves indicate that most of the service time for an average claim is spent in the first 18 months, which contrasts with the common belief that older claims require more service time per month to settle than those claims that are settled early and quickly.

The two most time-consuming activities of claims adjusting are the investigation of injuries to determine compensability and the coordi-

\textsuperscript{11}In most cases, the information needed to compile $F(t)$ is readily available from the TPA's computer system. On the other hand, the estimation of $g(t)$ may involve a detailed study on how claims examiners spend their time on claims with different ages. If such a study is not possible, $g(t)$ can be determined from the result of a survey based on examiners' experience and judgment.
nation of medical treatments that include surgeries and rehabilitation. Because these activities occur more frequently in the early stage of the claims, \( g(t) \) is usually a downward sloping curve or a humped curve for workers' compensation indemnity claims.\(^{12}\)

The next step is to determine the unit cost of examiner time (including salary, benefits, overhead, and profit) at the beginning of the contract period. For example, assume that the annual salary and benefits for an examiner are given at $50,000 while overhead and profit account for 50 percent of the cost. Given that the total working hours in a year are 2,000 (250 working days and eight hours per working day), the unit cost of examiner time can be set at $50 per hour \([(50,000/0.5)/2,000]\).

Let \( P^{(n)}(k) \) be the price for handling a new claim from month 1 to month \( k \) \((k = 1, 2, \ldots)\) and let \( c \) be the hourly cost of service time at the beginning of the contract period. Further assume that \( c \) increases at a rate of \((1 + s)\) per month, \( s \geq 0\). Thus, the hourly cost at the start of the \( t \)-th month is \( c_t = c \times (1 + s)^{t-1} \), for \( t = 1, 2, \ldots \). Then the per claim service price function for a new claim is given by:

\[
P^{(n)}(k) = \sum_{t=1}^{k} c_t \nu^{t-1} g(t)(1 - F(t)) \quad \text{for } k = 1, 2, \ldots \tag{1}
\]

where \( \nu = 1/(1 + i) \) is the monthly interest discount factor, with \( i \) being the monthly interest rate. The discount factor \( \nu \) can be selected by the TPA to reflect its cost of capital and other needs. We assume that all service time is rendered at the beginning of every month and, thus, discounting takes place at the beginning of each month, i.e., at time \( t - 1 \).

Equation (1) can be rewritten as

\[
P^{(n)}(k) = c \sum_{t=1}^{k} \beta^{t-1} g(t)(1 - F(t)) \tag{2}
\]

where \( \beta = (1 + s)\nu \). Thus the per claim service price for a new claim to be handled to settlement (for life)\(^{13}\) is \( P^{(n)}(\infty) \), while that for new claim service price for 12 month handling is \( P^{(n)}(12) \), and so on.

\(^{12}\)In establishing \( g(t) \), a TPA needs to consider segregating its experience into more homogeneous groupings. Experience may be subdivided by claim type or location (i.e., service time may be different as required by regulation. California and Texas are good examples.)

\(^{13}\)In this case, to ensure that \( P^{(n)}(\infty) \) is finite, we must have \( \beta < 1 \) or have the maximum number of years that it can take to settle a claim be bounded. In practice, the latter condition is not restrictive because one can expect all claims to be settled within say 30 years or 50 years or even 100 years.
Table 4 shows how service time and claim closure information are combined to develop the service fees for per claim. The cumulative closing percentage (\(F(t)\) in Column 4) at the beginning of the first month (\(t = 1\)) is zero. By the end of the month, 10 percent of the claims are closed and the service time rendered in the month is ten hours per claim. Thus, the expected service time for the first month is ten hours as indicated in the last column of the table. \(g(t)\) is the service time for each claim open at age \(t\). For the second month, \(g(2)\) is 14 hours and \((1 - F(2))\) is 90 percent. Therefore, the expected service time in the second month is 12.6 hours. It is straightforward to calculate the expected service time for the remaining months. Thus, for example, equation (2) leads to

\[ P^{(n)}(\infty) = c[10 + 12.6\beta + 11.4\beta^2 + \cdots + 0.78\beta^{11} + \cdots 0.1\beta^{23} + \cdots]. \]

<table>
<thead>
<tr>
<th>(t)</th>
<th>(g(t))</th>
<th>Closing %</th>
<th>(F(t))</th>
<th>(1 - F(t))</th>
<th>(g(t)(1 - F(t)))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10 hours</td>
<td>10%</td>
<td>0%</td>
<td>100%</td>
<td>10 hours</td>
</tr>
<tr>
<td>2</td>
<td>14 hours</td>
<td>14%</td>
<td>10%</td>
<td>90%</td>
<td>12.6 hours</td>
</tr>
<tr>
<td>3</td>
<td>15 hours</td>
<td>12%</td>
<td>24%</td>
<td>76%</td>
<td>11.4 hours</td>
</tr>
<tr>
<td>4</td>
<td>13 hours</td>
<td>11%</td>
<td>36%</td>
<td>64%</td>
<td>8.32 hours</td>
</tr>
<tr>
<td>5</td>
<td>10 hours</td>
<td>10%</td>
<td>47%</td>
<td>53%</td>
<td>5.3 hours</td>
</tr>
<tr>
<td>6</td>
<td>8 hours</td>
<td>9%</td>
<td>57%</td>
<td>43%</td>
<td>3.44 hours</td>
</tr>
<tr>
<td></td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
</tr>
<tr>
<td>12</td>
<td>3 hours</td>
<td>2.5%</td>
<td>74%</td>
<td>26%</td>
<td>0.78 hours</td>
</tr>
<tr>
<td></td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
</tr>
<tr>
<td>24</td>
<td>1 hour</td>
<td>1.2%</td>
<td>90%</td>
<td>10%</td>
<td>0.1 hour</td>
</tr>
<tr>
<td></td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
</tr>
</tbody>
</table>

For life of partnership handling, a subjective probability distribution has to be included to indicate the possibility of cancellation. In general, it is assumed that the average time of a contractual relationship between a self-insured and a TPA is three to five years. Consequently, the variations in life of partnership pricing among TPAs can be significant, depending critically on the expectation and the risk tolerance level of the TPA.
In establishing claim closure distributions, a TPA needs to consider segregating its experience into more homogeneous groupings. Long-tail lines usually exhibit distinctive closing patterns compared to other product lines. Even within a long-tail line it is usually beneficial to subdivide experience by claim type. For example, in workers' compensation, most medical-only claims can be closed within six months while some indemnity claims can linger for more than five years.

There is no doubt that this procedure can establish only a baseline for pricing while much of the pricing decision has to be based on the underwriting characteristics of the customers. One needs to examine, among other things, the claim closing patterns of the prospective clients in order to determine the deviation of their experience from the TPA's own experience and adjust the price accordingly.

3.2.2 Open Claims

Let \( P^{(o)}(m, n) \) be the service fee for an open claim at age \( m \) to age \( n \), for \( m = 1, 2, \ldots \) and \( n = m + 1, m + 2, \ldots \) Using the same notations as in Section 3.2.1, \( P^{(o)}(m, n) \) can be calculated as follows:

\[
P^{(o)}(m, n) = \sum_{t=m}^{n} c_t v^{t-m} g(t) \frac{(1 - F(t))}{(1 - F(m))}
= c (1 + s)^{m-1} \sum_{t=m}^{n} \beta^{t-m} g(t) \frac{(1 - F(t))}{(1 - F(m))}.
\] (3)

In practice, service charges for claims open more than 12 months are seldom based on individual claim age, as it would be tedious to calculate the fees. A weighted-average charge is applied to each open claim regardless of its age. Assuming the claim volume from year to year is stable, the formula for the weighted average charges is:

\[
P^{(o)}_w(m, n) = c (1 + s)^{m-1} \sum_{j=12}^{\infty} w_j \sum_{t=m}^{n+j} \beta^{t-m} g(t) \frac{(1 - F(t))}{(1 - F(m))}.
\] (4)

where

\[
w_j = \frac{1 - F(j)}{\sum_{k=12}^{\infty} (1 - F(k))}
\]

is the probability weight used for the \( j \)-th month.

\footnote{There is no assumption of universal efficiency of all TPAs. The poor claim experience of a customer, for example, when compared to similar risks, may result from the random variations of claims, the poor management of its TPA, or simply reflect the customer's true exposures.}
3.3 State-Group Relativities for Per Claim

For a TPA with clients in multiple states, there is a need to differentiate service costs among states. To calculate per claim charges by state, one can establish state-group relativities similar to those used in class ratemaking in property/casualty insurance pricing. Once state-group relativities are established, updates of the base price for each state can be performed easily.

The criteria to divide states into state-groups with similar claims handling costs can be based on the TPA's internal claims closure experience and cost by state, supplemented by statistics from national or state rating bureaus. For workers' compensation, important statistics include the percentage of serious cases and the per claim severity which may differ significantly by state. In addition, the degree of state regulation which is always an important contributing factor to TPA's service costs can help determine the makeup of the state-groups.

Specific actuarial techniques and much more data are needed to establish credible estimates of the values of state-group relativities. Even a national TPA may not have enough information in all claim categories for all states. For local or regional TPAs, state-group relativities can be set only judgmentally based on the TPA's internal cost and published information from state rating bureaus.

4 Incentive Contracts

The last pricing issue to be discussed is the design of incentive contracts. There has been a strong interest among self-insureds to establish a relationship between service fees and TPA performance in order to monitor the effectiveness of TPAs in controlling claim costs. Essentially, an incentive program requires that a certain percentage of the service fees be set aside for a bonus or penalty based on several performance measurements of the TPA services. The results of the performance measurements valued as of predetermined dates are compared to negotiated targets for the calculation of the bonus or penalty.

Before discussing any specific performance measurements, it is useful to set some common sense criteria to evaluate their feasibility. The following provides a reasonable checklist for such purposes:

- The TPA has sufficient control over the performance measurement;
- The value of the performance measurement can be objectively determined, and both parties have the ability to track results; and
There exist reliable benchmark data for comparison.

4.1 Basics

There are two major types of performance measurements: activity-based measurements and financial measurements. Popular measurements of TPA performance are usually activity-based such as number of claims closed by age, timely bill payments, timely claim processing, and reserving adequacy. The usual financial measurements for incentive programs include paid loss and incurred loss.

Most activity-based measurements can satisfy the three criteria. Take timely bill payments and claim processing as examples. An incentive program can stipulate that claim bills should be paid by the TPA within two business days after receiving the bills, or that claimants should be contacted within 24 hours after the claim is reported. The data for calculating such performance measurements should be available from the TPA's system and the results of the measurements can be determined easily. Therefore, the implementation of such an activity-based incentive program is usually straightforward.

4.2 Financial Incentive Contracts

The TPA industry has been experiencing more demand for financially based measurements, such as comparing actual and target incurred/paid amounts for claims incurred within the service contract period. In general, TPAs are hesitant to accept financially based measurements as they may appear to be taking insurance risk in which they have insufficient knowledge and little interest. Given that financial-incentive contracts have gained considerable popularity in recent years, the TPA industry has been forced to develop measurements that are mutually agreeable to the claims administrator and the self-insured.

Total policy year paid or incurred loss by development age have been suggested as performance measures for a risk-sharing program. Paid or incurred loss by development age is measured against an index such as policy year payroll before it is compared to a predetermined goal. Using the criteria described at the beginning of this section, it is clear that the amount of paid or incurred loss by development age per se can be determined easily. The TPA does not have sufficient control over the measures, however, as any total losses are affected by frequency, exposure, inflation, and other factors. In addition to the volatility of paid and incurred losses, it is difficult to find reliable data for benchmarking. Although these drawbacks may seem obvious to casualty actuaries,
many self-insureds insist on using changes in paid-to-date or incurred-to-date loss as performance measurements.

4.3 A Suggestion: Use Averages

Take workers' compensation as an example. There are four factors that can significantly change the financial results of a self-insured program: exposure (payroll) changes, state benefit changes, claim frequency changes, and inflation. A TPA should not be responsible for variations due to changes in exposure, frequency, and benefit level because none of these factors can be controlled by a claims administrator. For example, higher frequency in reported workers' compensation claims can be the result of a lay-off, which is beyond the control of the TPA.

To eliminate the impact of frequency changes on total loss, it seems appropriate and equitable to use incurred per claim severity as a performance measure for a financial risk-sharing plan. By eliminating the variations in frequency and exposure, per claim severity usually exhibits stable development patterns, given sufficiently large claim volumes. Most importantly, per claim severity can be managed and partially controlled by the TPA. Thus, it appears to be an ideal candidate for measuring TPA performance.

Additional benefits of using per claim severity as a performance measure are:

- There is no need to compare per claim severity to payroll or number of employees for incentive contract purposes;
- The industry average cost per claim by state is available from state rating bureaus; consequently, benchmarking should be easier and the results should be much more reliable;
- By comparing to an industry average, the variations due to changes in benefit level can be eliminated.

Per claim severity should be used on an ultimate basis as a performance measurement. Only when the baseline for comparison is established on an ultimate basis can the loss experience of a policy year be evaluated. The results can be misleading if one is merely looking

---

15 Precautions must be taken when bureau data are used, as trends and development may be needed.
16 A method for estimating ultimate values needs to be agreed upon in advance to avoid competing estimates from the TPA and the self-insured.
for the incremental changes between two development ages that are subject to the timing of claim payments and reserve recording.

An incentive contract ideally can look and operate in a way similar to retrospective rating plans. To establish a baseline for a policy year, the usual actuarial methods including capping large losses can be applied to loss data in the estimation of the ultimate severity. This can be done six months after the end of the policy year, the same time when retrospective rating plans start to evaluate policy year experience. The main difference is that in retrospective rating the target incurred loss is revised every 12 months thereafter until the final settlement of the policy year, while in incentive contracts a baseline (i.e., estimated ultimate severity) is determined six months after the end of the policy year for benchmarking purposes at later dates. A bonus or penalty can be calculated based on the deviation of the projected ultimate per claim severity at a later evaluation date (e.g., 30 months after policy inception) from the baseline. A subsequent computation/adjustment can be performed every 12 months until both parties agree that the latest computation will be the final one for the policy year.

5 Concluding Remarks

One important component that is missing in TPA pricing is self-insurance database support. Self-insured entities do not report loss, payroll, or other relevant experience data to state rating bureaus. To meet their pricing needs, TPAs rely on their own experience or purchase data from state rating bureaus, which may or may not be appropriate for the self-insurance purposes. The National Council on Compensation Insurance has initiated a program for collecting loss data on self-insured groups. This may be a good start toward a more complete and reliable database for TPA pricing.

With the introduction of managed care organizations (MCOs) in many states, the role of TPAs in the business of claims handling may fundamentally change. Judging from developments over the past few years, TPAs and MCOs may have to share the responsibilities in medical cost containment, rehabilitation, and return-to-work programs. On the other hand, TPAs may be in an excellent position to launch their own medical networks and merge these two functions. It will be interesting to see how these changes will impact the pricing of traditional TPA services and the expanded services provided jointly by a TPA and an MCO.
References

