JOURNAL OF HEALTH CARE FINANCE

Fall 2017

Correlating Medicaid Reimbursement Reductions, the Sustainability of Providers, and Access to Care Utilizing Healthcare KPIs

Alfredo Sepulveda, PhD
Principal Researcher, QDR Labs

Abstract

Do reimbursement rates to Medicaid providers correlate with their Medicaid participation rates. That is, will Medicaid reimbursement rates dictate trends in provider Medicaid participation because of fiscal stability in their payor portfolios? The implementation of the Medicaid Primary Care Payment Increase as part of the initial ACA rollout proved to be a murky testing ground for this thesis. In part, the short-term increase coupled with states' delayed implementation retarded any large effects of such increases. Nonetheless, the consensus was positive, and the increase prompted more enthusiasm to participate in Medicaid programs and increase access to care (RAND Corporation, 2017). In states opting into Medicaid expansion in the Affordable Care Act (ACA) program, the answer was initially not clear (MACPAC, 2018). Nonetheless, Medicaid provider participation has not seen a systematic national drop – it has stabilized around an average 70% mark for several years into the ACA implementation. However, states' Medicaid reimbursement rates differ and there are microeconomic correlations pointing to the effect of lower Medicaid reimbursements in state regions. This paper will focus on microeconomic key performance indicators (KPIs), that may empirically dictate the effect of reimbursement rate reductions in Medicaid on provider business stability and patient access to care epiphenomena.

Introduction

The recent global and national economic malaise that has produced budgetary belt-tightening in the legislative houses of states and Washington, D.C. has also made healthcare costs the poster boy for implementing financial hyper-efficiency. Value-based and hybrid methodologies thereof, have been proposed and implemented by CMS for Medicare reimbursements in the ACA system of Accountable Care Organizations (ACOs) and by many private healthcare payor programs, in an attempt to redefine provider reimbursement efficiency. However, Medicaid reimbursement remains based on a fee-for-service (FFS) methodology. Nonetheless, pronounced reductions have been made to states' Medicaid programs, in addition to those programs being converted to managed care methodologies and authority. State budget shortfalls, the unpredictability of the passage of a version of the American Health Care Act (AHCA) in the future as a change to certain ACA features or repeal altogether, and decreased federal matching funds for Medicaid programs have promulgated this trend. Additionally, it appears that the states' Children's Health Insurance Program (CHIP) programs, insuring children and pregnant women whose families are marginally above the federal poverty level (normally between 200% and 250% of FPL) are in peril because of, at best, a short-term stop-gap federal spending bill in 2017. States normally cut Medicaid subprograms and CHIP spending through lower provider reimbursements, medical authorization reform, stricter eligibility requirements, and more recently, implementation of patient co-payments in Medicaid. The potential for decreased Medicaid provider participation because of these fiscal changes, is of concern because of the *presumed* access to care shrinkage consequent to these changes and to the growing Medicaid population. The ACA rollout also introduced further growth in the Medicaid population with expanded Medicaid coverage to adults in states opting into that subprogram.

The implementation of the Medicaid Primary Care Payment Increase as part of the initial ACA rollout was done to stimulate growth of provider participation in Medicaid programs and increase patient access to care in those programs, in anticipation of increased Medicaid rolls in states. This proved to be difficult to implement universally nationwide. States' fiscal limitations prevented smooth transitions to payer systems and the effect of these reimbursement increases was dulled. There were also political implications to the rollout in some states opting out of the Medicaid expansion in the ACA. In the whole, providers were encouraged to participate more in Medicaid programs but were reluctant to commit long-term because of the short two-year span of the increases (RAND Corporation, 2017).

While these increases were implemented, states anticipated shortages in their healthcare budgets. This meant that those national increases were lessened considering state-contributed reimbursement reductions. Most Medicaid non-primary care providers experienced massive rate reductions in most states with many being reimbursed at markedly sub-Medicare rates. Medicaid provider reimbursement reductions continue to be implemented, along with stricter authorization protocols. This phenomenon has put a greater strain on the Medicaid provider to continue servicing the Medicaid population or worse still, stay afloat in Medicaid dense areas of practice. This trend, in turn, creates access to care gaps for many specialties providing services to the Medicaid population.

In this short treatise, microeconomic financial key performance indicators (KPIs) are utilized to formulate the status of a healthcare provider's viability in relation to reimbursement rate reductions that have become entangled in the budget quagmire of government healthcare economics. The microeconomic financial KPIs of (i) gross profit margin, (ii) return-on-investment, and (iii) Tobin's Q measurement are used to equate a reimbursement rate reduction to the ongoing worth of the healthcare provider. Micro healthcare financial KPIs such as claims denial rates and processing costs, operational and marketing costs, and administration and clinical staff costs will be wrapped into these macro KPIs. Patient accessibility is correlated with these measures (both macro and micro) and thus with future reimbursement rate reductions. Since government reimbursement rates are directly linked to all healthcare insurance portfolios, these rates are universal indicators for the performance and sustainability of the healthcare provider.

Certainly, other micro healthcare operational KPIs are relevant in estimating the worth and efficiency of a healthcare practice. These include (i) patient wait times, (ii) number of treatment rooms in use at any one time, (iii) staff-to-patient ratio, (iv) treatment room turnover, (v) general communication flow and information accuracy between the healthcare coordinated team and the patient, (vi) patient discharge, cure, and recidivism rates, (vii) patient satisfaction rates, (viii) provider diagnostic, medication, and treatment error rates, (ix) patient follow-up rates, (x) average treatment times, (xi) cancelation and no-show (DNA) rates, (xii) use of evidence-based methods and practices, and most importantly, (xiii) patient outcomes. These measures are more relevant in estimating value-based reimbursements. However, they contribute to the overall macro financial KPI of overhead costs per treatment service. Reimbursement rate schedules can be expressed as functions of such micro KPIs as in value-based purchasing (VBP) methodologies and carveouts. Hence, the subject here, remains applicable to hybrids and mixtures of traditional reimbursement methodologies such as VBP.

Gross profit margins, return-on-investment (ROI), and Tobin's Q measurement will be initially reviewed. These indicators will then be formulated as functions of issued reimbursement rate schedule reductions as percentages of current rates. Patient accessibility as a measure of a viable healthcare firm are then expressed in terms of these indicators and to reimbursement rate schedule reductions.

Healthcare Key Performance Indicators and Reimbursements

Revenue-to-cost metrics are standard absolute measures of a firm's health. However, they may also be misleading in the fact that larger firms may have relatively larger revenues but still suffer from underwhelming profits. Small firms, on the other hand, may have lower relative revenues, but have sustainable profits. To compare healthcare firms equitably, we use relative measures of profitability. In this vein variants of profit margins are used to normalize profit measurement. Consider the gross profit margin of a firm. The gross profit margin (*gpm*) is the relative unit of profit per unit of revenue, considering the cumulative relative unit of cost:

$$gpm = \frac{r-c}{r} = 1 - \frac{c}{r} \tag{0.1}$$

where *r* is unit revenue and *c* is unit cost. Gross profit margins are relative measures of sustainability and are not directly dependent on the size of the firm. If *c* represents the total costs (direct and indirect) related to servicing the revenue generator for *r*, then *gpm* is a *net profit margin* measurement. We will assume that all *gpm* measurements will be net profit margin (*npm*) measurements as they are more accurate measurements of performance. Gross margins $gm = r - c$, are more often used in healthcare finance economics when absolute metrics are more important for calculating absolute per unit performance (Gapenski & Reiter, 2016). However, we are more interested in the overall sustainability of a healthcare firm and as such will utilize the net gross profit margin as a starting metric. For any given treatment service, *i*, let

$$gpm_i = 1 - \frac{r_i}{c_i} \tag{0.2}$$

denote the net profit margin for that service treatment item. These services represent the potential revenue streams for the healthcare provider. Denote each unit revenue stream amount by r_i. Now consider the total costs to the healthcare provider in preparing and rendering such a unit service, c_i. One can calculate net profit margins by service type or cumulatively by the sum of their revenue and cost streams.

Each reimbursable revenue stream is further divided into prospective payor types such as government programs, (i.e., Medicaid, Medicare, Foster Care, etc.), private insurance, and cash payment. With managed care programs, government reimbursements are managed by private insurance where those insurance carriers set their own fee schedules that are actuarially dependent on the government reimbursement schedules, (i.e., state Medicaid fee schedules or the CMS Medicare Physicians Fee Schedule Resource-based Relative Value Scale-MPFS RBRVS). In preparing for rate reductions, (i.e., reductions in reimbursement fee schedules), prospective fee schedules can be compared to baseline current or past fee schedules through rate reduction percentages, to be denoted here by α. These relative reductions will directly impact profit margins and all measures of profitability and sustainability for healthcare firms that depend on any form of reimbursement, including cash payment because the CMS MPFS directly influences what the standard harmonized cost of healthcare services will be, (i.e., all schedules are compared directly to the MPFS RBRVS).

Cost can be divided into direct and indirect costs associated with a service item, $c_i = c_i^d + c_i^{id}$. These cost components can in turn be broken down by categories such as medical device and tools usage, utilities, staff time and wages, real estate (mortgage and/or lease), computer usage, administrative overhead, billing and collections, etc. Within these categories, subcategories can be used to differentiate fixed versus variable costs. Variable costs can be calculated via volume of units and unit prices. However, there is always stochasticity involved in volume usage and unit prices, dependent on economic cycles such as the Personal Consumption Expenditures chain-type price index (PCECTPI)

In order to more accurately calculate such an impact on profits, the rate reductions (either individual service code, weighted averages, or cumulative) can be used to express profit margins in terms of a function of the rate reductions, α. For a given service code, *i*, denote the individual rate reduction for that service by α_i. We differentiate between average and weighted average rate reductions. The weights used are revenue based. One would need to consider an effective rate reduction by implementing revenue-based weights. To this end, consider for service code *i*, the frequency of service, f_i and the current reimbursement amount, R_i. The *effective rate reduction,* $\bar{\alpha}_w$ expressed as the revenue-based weighted average rate reduction is:

$$\bar{\alpha}_w = \frac{\sum_{i \in P} f_i R_i \alpha_i}{\sum_{i \in P} f_i R_i} \tag{0.3}$$

where *P* is the index of potential service treatments rendered by the healthcare provider and $w = \left(w_i\right)_i = \left(f_i R_i\right)_i$ is the weight vector. We now define the net profit margin as a function of the rate reduction α_i:

$$gpm_i\left(\alpha_i\right) = 1 - \frac{c_i}{(1-\alpha_i) r_i} \tag{0.4}$$

given the total cost, c_i and revenue r_i for such a treatment service. To calculate the delta or difference in profit margins that has occurred as a result of such rate reductions, express the difference between the old and new profit margins,

$$\begin{aligned} \delta_i\left(\alpha_i\right) &= \Delta gpm_i\left(\alpha_i\right) = gpm_i - gpm_i(\alpha_i) \\ &= c_i\left(\frac{1}{(1-\alpha_i) r_i} - \frac{1}{r_i}\right) \\ &= \frac{c_i}{r_i}\left(\frac{\alpha_i}{1-\alpha_i}\right) \\ &= \beta_i\left(\frac{\alpha_i}{1-\alpha_i}\right) \end{aligned} \tag{0.5}$$

where $\beta_i = \dfrac{c_i}{r_i}$ is the usual cost-to-revenue ratio for service item *i*.

Individual rate reduction percentages, α , are realistically in the range, $0 < \alpha < 1$, and are typically incrementally limited by political, economic, and strategic forces within the government and private industry coalitions to the range, $0 < \alpha < 0.5$. $\delta_i\left(\alpha_i\right)$ is a hyperbolic function in α_i , ranging from 0 at $\alpha_i = 0$ (no rate change), to asymptotically approaching ∞ as $\alpha_i \rightarrow {}^{-}1$ (100% reduction, no reimbursement). However, unsustainability of the firm usually occurs well before this descent, typically when $\dfrac{\beta_i}{\left(1-\alpha_i\right)} \approx 1$ (the new cost-to-revenue ratio approaches unity), or as $\delta_i\left(\alpha_i\right) \rightarrow \alpha_i$, (i.e., the gross profit margin differential approaches the rate reduction). At $\alpha_i = 0.5$ (50% reduction), $\Delta gpm_i\left(0.5\right) = \beta$ and the gross profit margin has been reduced by one factor of the firm's previous cost-to-revenue ratio. Minimally, in a sustainable firm, $\beta < 1$. In practice, however, a threshold, $\delta > 0$, is reached (tipping point) for unsustainability when $1 > \Delta = \left|\beta - 1\right| > \delta > 0$.

These gross profit margin calculations are with respect to one service item, that is, they measure a relative profitability of rendering a service item, given a rate reduction for that service item. To calculate a cumulative gross profit margin or total gross profit margin one needs to accumulate an estimate of the revenues from all service items. Since gross profit margins are calculated with respect to a time period, average frequencies of servicing those service items can be used for a time period, typically monthly. Additionally, reimbursements types, (i.e., payor fee schedules) and their respective frequencies should be used for a more accurate revenue portfolio. Finally, since all fee schedules are calculated with respect to the RBRVS (resource-based relative value scale) used in the MPFS, the amounts should be expressed as such, (i.e. as percentages of the RBRVS values for each service item). The usual and customary fee schedule for providers (U&C) is usually calculated as percentages of RBRVS. However, any of the amounts noted in the calculation of profit margins or any other component used in a measure of sustainability can be expressed as a percentage of RBRVS. In this way, the calculations are standardized to the RBRVS and direct comparisons can be made across and between payor and provider fee schedules.

We now express the gross profit margin with respect to a payor portfolio:

$$gpm(p) = 1 - \frac{\sum_{i \in P} f_{p,i} c_i}{\sum_{i \in P} f_{p,i} r_{p,i}} = 1 - \frac{\sum_{i \in P} f_{p,i} k_i R_i}{\sum_{i \in P} f_{p,i} s_{p,i} R_i} \qquad (0.6)$$

where $f_{p,i}$ is the frequency of service *i* to patients in the practice who are network members of payor *p*,

R_i is the MPFS RBRVS reimbursement for service *i*,

$k_i = \dfrac{c_i}{R_i}$, is the % of RBRVS of the cost, c_i , of *i*,

$s_{p,i} = \frac{r_{p,i}}{R_i}$, is the % of RBRVS of the reimbursement, $r_{p,i}$, from payor p,

and P is the list of services reimbursed by payor p to practice.

When a rate reduction is introduced, this initializes an avalanche effect upon all payor schedules. Hence, we must calculate the aggregate of (0.6) summed over all payors in a provider payor portfolio, *U*.

$$gpm(U) = 1 - \frac{\sum_{p \in U} \sum_{i \in I_p} f_{p,i} c_i}{\sum_{p \in U} \sum_{i \in I_p} f_{p,i} r_{p,i}} = 1 - \frac{\sum_{p \in U} \sum_{i \in I_p} f_{p,i} k_i R_i}{\sum_{p \in U} \sum_{i \in I_p} f_{p,i} s_{p,i} R_i} \tag{0.7}$$

If α_i is the rate reduction for service item *i*, then we have the new baseline rate expressed as:

$$R_i(\alpha_i) = (1 - \alpha_i) R_i \tag{0.8}$$

and

$$s_{p,i}(\alpha_i) = \frac{r_{p,i}(\alpha_i)}{R_i(\alpha_i)} = \frac{\tau_{i,p} r_{p,i}}{(1-\alpha_i) R_i} = \frac{\tau_{i,p}}{(1-\alpha_i)} s_{p,i} = \mu_{i,p} s_{p,i} \tag{0.9}$$

Here $\tau_{i,p}$ is the adjusted payor reduction rate as a result of the reduced baseline rate α_i and

$\mu_{i,p} = \frac{\tau_{i,p}}{(1-\alpha_i)}$ is what we call the *effective rate reduction from payor p*. Assume that $\tau_{i,p}$ is functionally actuarially dependent on α_i, (i.e., $\tau_{i,p} = \tau_{i,p}(\alpha_i)$) and so, $\mu_{i,p} = \mu_{i,p}(\alpha_i)$ is. If we substitute (0.8) and (0.9) into (0.6), we obtain the rate reduced gross profit margin with respect to payor *p*:

$$gpm(p)(\alpha) = 1 - \frac{\sum_{i \in P} f_{p,i} \frac{k_i}{(1-\alpha_i)} R_i(\alpha_i)}{\sum_{i \in P} f_{p,i} \mu_{i,p}(\alpha_i) s_{p,i} R_i(\alpha_i)} \tag{0.10}$$

where $\alpha = (\alpha_i)_{i \in P}$ is the array of rate reductions applicable to the fee schedule of payor *p*. Expression (0.10) can be rewritten as:

$$gpm(p)(\alpha) = 1 - \frac{\sum_{i \in P} c_{p,i}(\alpha_i) R_i(\alpha_i)}{\sum_{i \in P} r_{p,i}(\alpha_i) R_i(\alpha_i)} \tag{0.11}$$

where we label the new weights, $c_{p,i}(\alpha_i) = f_{p,i} \frac{k_i}{(1-\alpha_i)}$ and $r_{p,i}(\alpha_i) = f_{p,i} \mu_{i,p}(\alpha_i) s_{p,i}$ as the *cost* and *revenue payor weights* respectively with respect to the baseline rate reduction α_i.

Population Dynamics and Carve Outs

In (0.10) the service code frequencies, $f_{p,i}$ are dependent on the payor network insuree population in the provider's service area. Using historical facility claims data, one may estimate $f_{p,i}$ by using relative frequencies based on usage statistics for service *i* in the service area during a given time period. Seasonal averages may be harmonized in order to obtain a monthly average frequency. One can also proceed as follows. First, obtain an estimate of the percentage of eligible insures that will need service *i* during the reporting period. Let n_p be the total number of eligible insurees for payor *p* in the service area, $q_{p,i}$ be the percentage of insurees in service area that will need service *i*, $a_{p,i}$ be the average number of visits during the reporting period for service *i* for payor insuree population, and $ms_{p,i}$ be the market share in service area of payor insure population for service *i*. $q_{p,i}$ and $a_{p,i}$ can be estimated based on larger regional or state admittances using Medicare or Medicaid historical population statistics or by using historical facility data. $ms_{p,i}$ can be estimated based on competitor population in service area, (i.e., assuming a uniform spread and letting nc_p be the number of competitors in the service area contracted with payor *p*, then $ms_{p,i} \approx \frac{1}{nc_p}$). See the Appendix for a measure of market share based on divergence measures between prospective insurees and a facility. Now estimate the period service frequency as:

$$\hat{f}_{p,i} = a_{p,i} q_{p,i} n_p ms_{p,i}\,, \tag{0.12}$$

using this to do the subsequent estimate for the weights:

$\hat{c}_{p,i}(\alpha_i) = a_{p,i} q_{p,i} n_p ms_{p,i} \frac{k_i}{(1-\alpha_i)}$ and $\hat{r}_{p,i}(\alpha_i) = a_{p,i} q_{p,i} n_p ms_{p,i} \mu_{i,p}(\alpha_i) s_{p,i}$ and finally obtaining an estimate for the rate reduced gross profit margin:

$$gpm(p)(\alpha) = 1 - \frac{\sum_{i \in P} \hat{c}_{p,i}(\alpha_i) R_i(\alpha_i)}{\sum_{i \in P} \hat{r}_{p,i}(\alpha_i) R_i(\alpha_i)} \tag{0.13}$$

One may also do individual *carve-out* analysis on a payor reimbursement schedule using:

$$gpm(p)(\alpha_i) = 1 - \frac{\hat{c}_{p,i}(\alpha_i) R_i(\alpha_i)}{\hat{r}_{p,i}(\alpha_i) R_i(\alpha_i)} \tag{0.14}$$

$gpm(p)(\alpha_i)$ would then be an indicator of the economic viability and profitability of contracting with payor *p* with a carve-out that would include service item *i*. $gpm(p)(\alpha)$ would be a similar indicator for the reimbursement fee schedule *P* from payor *p*. To consider a carve-out option, let $I_{O_j,p} = \left\{ i^1_{O_j,p}, i^2_{O_j,p}, \ldots, i^{n_{O_j,p}}_{O_j,p} \right\}$ denote the indices of service codes that the provider O_j, will emphasis from payor *p* fee schedule. The space of provider competitors, $O=\{O_1, O_2, ..., O_M\}$ is limited to those healthcare firms that will economically impact each other in the larger market of healthcare firms. Then the *gpm* for provider O_j, over payor *p* fee schedule, can be separated into two distinct components:

$$gpm(O_j, p)(\alpha) = 1 - \left[\frac{\sum_{i \in P \setminus I_{O_j,p}} \hat{c}_{p,i}(\alpha_i) R_i(\alpha_i) + \sum_{i \in I_{O_j,p}} \hat{c}_{p,i}(\alpha_i) R_i(\alpha_i)}{\sum_{i \in P \setminus I_{O_j,p}} \hat{r}_{p,i}(\alpha_i) R_i(\alpha_i) + \sum_{i \in I_{O_j,p}} \hat{r}_{p,i}(\alpha_i) R_i(\alpha_i)} \right] \quad (0.15)$$

In the carve-out profile for O_j, the deemphasized service codes, $P \setminus I_{O_j,p}$ have relatively small service frequencies. So, the emphasized costs and revenues overwhelm the deemphasized in (0.15), giving the approximation:

$$gpm(O_j, p)(\alpha) \approx 1 - \frac{\sum_{i \in I_{O_j,p}} \hat{c}_{p,i}(\alpha_i) R_i(\alpha_i)}{\sum_{i \in I_{O_j,p}} \hat{r}_{p,i}(\alpha_i) R_i(\alpha_i)} \quad (0.16)$$

In a legally negotiated carve-out proposal, a messenger model-based approach can be used in which one facility at a time may negotiate with a payor for one contract, without any possible collusion among a subgroup of facilities. In the case of an entire industry, cumulative frequencies can be used to calculate service code weights and be incorporated into the industry carve-out proposal to government entities.

Now consider the more adaptive case in which the facility adapts to the rate reduction schedule by adjusting its cost structure, (i.e., c_i is a function of $\alpha_i, c_i(\alpha_i)$). The gross profit margin expression (0.4) changes as:

$$gpm_i(\alpha_i) = 1 - \frac{c_i(\alpha_i)}{(1-\alpha_i) r_i} \quad (0.17)$$

In the case of a simple percentage cost reduction, (0.17) can be generally expressed as:

$$gpm_i(\alpha_i) = 1 - \frac{(1 - d_i(\alpha_i)) c_i}{(1-\alpha_i) r_i} \quad (0.18)$$

where $d_i(\alpha_i)$ is a function of the rate reduction α_i. Expression (0.13) can then be rewritten as:

$$gpm(p)(\alpha) = 1 - \frac{\sum_{i \in P} c^*_{p,i}(\alpha_i) R_i(\alpha_i)}{\sum_{i \in P} r_{p,i}(\alpha_i) R_i(\alpha_i)} \quad (0.19)$$

where $c^*_{p,i}(\alpha_i) = f_{p,i} \frac{k_i (1 - d_i(\alpha_i))}{(1-\alpha_i)}$. Expression (0.13) becomes:

$$gpm^*(p)(\alpha) = 1 - \frac{\sum_{i \in P} \hat{c}^*_{p,i}(\alpha_i) R_i(\alpha_i)}{\sum_{i \in P} \hat{r}^*_{p,i}(\alpha_i) R_i(\alpha_i)} \tag{0.20}$$

where $\hat{c}^*_{p,i}(\alpha_i) = a_{p,i} q_{p,i} n_p ms_{p,i} \frac{k_i\left(1 - d\left(\alpha_i\right)\right)}{(1-\alpha_i)}$ and $\hat{r}^*_{p,i}(\alpha_i) = a_{p,i} q_{p,i} n_p ms_{p,i} \mu_{i,p}(\alpha_i) s_{p,i}$ are the new cost and revenue payor weights respectively. In a carve-out option, (0.20) is expressed as:

$$gpm^*_c(O_j, p)(\alpha) \approx 1 - \frac{\sum_{i \in I_{O_j,p}} \hat{c}^*_{p,i}(\alpha_i) R_i(\alpha_i)}{\sum_{i \in I_{O_j,p}} \hat{r}^*_{p,i}(\alpha_i) R_i(\alpha_i)} \tag{0.21}$$

Summing (0.21) over a provider's payor portfolio, *U*, one obtains the aggregate carve out *gpm*:

$$gpm^*_c(O_j, U)(\alpha) \approx 1 - \frac{\sum_{p \in U} \sum_{i \in I_{O_j,p}} \hat{c}^*_{p,i}(\alpha_i) R_i(\alpha_i)}{\sum_{p \in U} \sum_{i \in I_{O_j,p}} \hat{r}^*_{p,i}(\alpha_i) R_i(\alpha_i)} \tag{0.22}$$

The aggregate differential between the current *gpm* and the rate reduced *gpm* would then be expressible as:

$$\delta^*_c(O_j, U)(\alpha) = \frac{\sum_{p \in U} \sum_{i \in I_{O_j,p}} \hat{c}^*_{p,i}(\alpha_i) R_i(\alpha_i)}{\sum_{p \in U} \sum_{i \in I_{O_j,p}} \hat{r}^*_{p,i}(\alpha_i) R_i(\alpha_i)} - \frac{\sum_{p \in U} \sum_{i \in I_{O_j,p}} f^*_{p,i}(\alpha_i) r^*_{p,i}(\alpha_i) R_i(\alpha_i)}{\sum_{p \in U} \sum_{i \in I_{O_j,p}} f^*{}_{p,i}(\alpha_i) s_{p,i}(\alpha_i) R_i(\alpha_i)} \tag{0.23}$$

where $r^*_{p,i}(\alpha_i) = f^*_{p,i}(1 - \mu_{i,p}(\alpha_i)) s_{p,i}$, $\mu_{i,p,O_j}(\alpha_i) = \frac{\tau_{p,i,O_j}(\alpha_i) - \alpha_i}{1 - \alpha_i}$ and $\tau_{p,i,O_j}(\alpha_i)$ is the (adjusted) payor *p* discount rate reduction for service item *i* for provider O_j, as a result of α_i. $\tau_{p,i,O_j}(\alpha_i)$ is usually compounded with α_i producing a final adjusted effective rate reduction of $1 - \tau_{p,i,O_j}(\alpha_i)[1 - \alpha_i]$.

Using ROI and Tobin's Q as Leading KPIs

Another measure of viability is the return on investment (ROI) in performing service *i* using the portfolio from payor *p*. ROI is more difficult to quantify because it involves, in addition to revenue, any future perception of investment or gain from performing that service. Such future investment may include an enhanced reputation for delivering quality care as recommendations from patients on your behalf to referring physicians, specialists, payors, and other patients. Also using outcome statistics, if the number of favorable outcomes or goals met as a percentage of total episodes is high or is recovery rate periods and maintainability of functionalities is high, the firm's ROI may be increased with respect to the service being rendered. Evidence-based outcomes are instrumental in measuring viable service outcomes for positive ROI. ROI is normally expressed as:

$$ROI_i = \frac{g_i - c_i}{c_i} = \frac{r_i + g_i^o - c_i}{c_i} = \frac{r_i}{c_i} - 1 + \frac{g_i^o}{c_i} = \gamma_i - 1 + \delta_i \qquad (0.24)$$

where g_i^o is the gain made from servicing service *i* from non-revenue directly, c_i is the cost and g_i is the total gain. γ_i is the usual revenue-to-cost ratio and κ_i is the less tangible non-revenue gain-to-cost ratio. $\kappa_i < 0$ if reputation is damaged due to some negligence, oversight, incompetence, or accident. So gain may be turned into a loss in these situations and a risk assessment is warranted. $ROI_i < 0$ is unsustainable unless other *ROI*s positively overwhelm it and it is part of the required complete service portfolio. *ROI* and gross profit margin are related as:

$$gpm_i = \frac{ROI_i - \kappa_i}{ROI_i + 1 - \kappa_i} \qquad (0.25)$$

The inverse relationship is:

$$ROI_i = \frac{gpm_i}{1 - gpm_i} + \kappa_i \qquad (0.26)$$

The carveout aggregate *ROI* with respect to a rate reduction schedule α, for provider O_j, with payor portfolio *U* would then be:

$$ROI_c^*(O_j, U)(\alpha) = \frac{gpm_c^*(O_j, U)(\alpha)}{1 - gpm_c^*(O_j, U)(\alpha)} + \kappa(\alpha) \qquad (0.27)$$

where $\kappa(\alpha)$ is the aggregate gain-to-cost ratio. Another measure of viability for a general firm is Tobin's *Q* measure given as:

$$Q_T = \frac{m_p}{a_p} \qquad (0.28)$$

where m_p is the perceived market price or value of the firm and a_p is the asset price of the firm which is normally assessed from a firm's balance sheet (Tobin & Brainard, 1977). m_p measures a market's perception of the firm's worth that would also include any intangible assets such as intellectual, strategic, and political capital and resources. The balanced scorecard methodology endeavors to quantify these values (Kaplan, & Norton, 1992; Sveiby, 2010). Nonetheless, there may be a negative correlation between Q_T and capital investment (Nitzan & Bichler, 2009). Q_T is also related to the $\frac{P}{B}$ ratio (price-to-book ratio). In an inflationary economy, $\frac{P}{B} > Q_T$, otherwise, traditionally, $Q_T \geq \frac{P}{B}$. In a purely rational market with no micro-fluctuations, $Q_T = \frac{P}{B}$.

It is posited that there may be a point of diminishing returns when further investment may actually decrease the market's perception of the firm's worth. To see this, a marginal version of Q_T may be expressed as the incremental ratio of a unit of capital investment to asset unit, $mQ_T = \frac{\Delta C}{\Delta A}$, where ΔC is a unit of capital investment and ΔA is a corresponding unit of asset worth. If $\Delta C = F(\Delta A)$ for some decreasing *perception function F*, of the asset capitalized, ΔA, then $mQ_T \to 0$, as $C \to \infty$. In this case, $\Delta C = F(\Delta A) = o(\Delta A)$. $Q_T > 1$ indicates that the market perception of the firm is hyper-inflated. In the case of $Q_T >> 1$, a valuation bubble exists. If a healthcare firm possesses an inflated Q_T or mQ_T value, it is being perceived by the market as having a strategic advantage within the reimbursement portfolio of government entities and healthcare insurance contracts, sustainable patient and provider referral censuses, and a stable service offering, (i.e., staff stability and service diversity). The asset price can be estimated based on the healthcare firm's tangible resources (equipment, staff expertise and specialties, location and strategic advantages in the market, etc.) and on future perceptions for profitability and revenue generation usually based on periodic frequencies of services rendered, its patient census, its insurance portfolio reimbursement fee schedules, and its referral network. We will later express Q_T as an approximate function of *gpm*.

Access to Care Differentials and Treatment Queuing

How would one equate measures of profitability, such as *gpm* to patient-centric issues, such as accessibility? One way would be to equate *gpm* to sustainment and enhancement of patient accessibility to healthcare and continuity of care. One could approximate what percentage of your positive *gpm* will be used to re-invest into your programs to sustain/enhance patient accessibility to care, retention, compliance and attendance and to have full access to your regiment of healthcare services.

No increase in your *gpm* will result if your new investment, given by *z*, and the resultant increase in revenues, *b* from that investment, are related as:

$$\frac{b}{z} \le \frac{R}{C} \tag{0.29}$$

where *R* is your current total revenue and *C* is your current total costs. That is, you are not enhancing your gross profit margin while simultaneously increasing or maintaining patient accessibility and continuity of care.

To retain accessibility for your patient census, while keeping profit margins flat or near flat after a rate reduction, the adjusted costs, $c_{p,i}(\alpha_i)$, must be increased. The relationship must approximately follow as:

$$\sum_{p \in U} \sum_{i \in I_{O,p}} c_{p,i}(\alpha_i) R_i(\alpha_i) \approx \sum_{p \in U} \left[\frac{\sum_{i \in I_{O,p}} r_{p,i}(\alpha_i) R_i(\alpha_i)}{\sum_{i \in I_{O,p}} r_{p,i} R_i} \right] \sum_{i \in I_{O,p}} c_{p,i} R_i \tag{0.30}$$

Let $z=\sum_{p\notin U}\left[\sum_{i\in I_{O,p}} c_{p,i}(\alpha_i)R_i(\alpha_i)-\sum_{i\in I_{O,p}} c_{p,i}R_i(\alpha_i)\right]$ and $b=\sum_{p\in U}\left[\sum_{i\in I_{O,p}} r_{p,i}(\alpha_i)R_i(\alpha_i)-\sum_{i\in I_{O,p}} r_{p,i}R_i\right]$ denote the added investment to retain patient accessibility and the added revenue respectively after a rate reduction, $\alpha=(\alpha_i)_{i\in I_{O,p}}$ has been implemented. Then by (0.30), the condition (0.29) is satisfied if:

$$\frac{b}{z}\underset{\sim}{<}\frac{\sum_{p\in U}\sum_{i\in I_{O,p}} r_{p,i}R_i}{\sum_{p\in U}\sum_{i\in I_{O,p}} c_{p,i}R_i} \tag{0.31}$$

Relative-value based methodologies are used to calculate reimbursement schedules such as the CMS MPFS RBRVS and the adoption of their relative values in some states' Medicaid fee schedules such as that in Texas known as the Texas Medicaid Fee Schedule (TMFS). The MPFS RBRVS system is based on the following 3-component formula:

$$r_i=\left[rvu_{i,w}gpci_{l,w}+rvu_{i,pe}gpci_{l,pe}+rvu_{i,mp}gpci_{l,mp}\right]CF \tag{0.32}$$

where *CF* is a universal dollar cost factor, the *rvu*s are relative-value units for each component involving work practice expense, and malpractice, with corresponding geographic practice cost indices, *gpci*. *i* is the index for the CPT based service. In Texas, the TMFS uses a reimbursement methodology in which only the RVUs are utilized and converts the formula using conversion units based on accessibility statistics, historic payment patterns, and deficiency-adjusted non-geographic weights, *a*, as:

$$r_i=\left[rvu_{i,w}a_{l,w}+rvu_{i,pe}a_{l,pe}+rvu_{i,mp}a_{l,mp}\right]a_{CF}CF \tag{0.33}$$

For the MPFS, both facility and non-facility rates are calculated using different *RVU*s. It can be proposed to use this same two-component system of facility and non-facility rates in calculating the TMFS as:

$$r_i=r_{i,f}+r_{i,nf} \tag{0.34}$$

where $r_{i,f}$ would encompass the facility (building) component of the service rate reimbursement, while $r_{i,nf}$ would include all technical components of the service. Rate reduction proposals should include adjustments made for facility costs, (i.e., $r_{i,f}$). While this is true of Medicare reductions, state Medicaid reductions do not usually include a facility cost component adjustment. Separate rate reductions could have be proposed for each component. In this way, (0.33) could be expressed in terms of a 2-component rate reduction, $\alpha=(\alpha_f,\alpha_{nf})$:

$$r_i(\alpha_i)=\begin{bmatrix}\left(1-\alpha_i^f\right)\left(rvu_{i,w}^f a_{i,w}^f+rvu_{i,pe}^f a_{i,pe}^f+rvu_{i,mp}^f a_{i,mp}^f\right)+\\ \left(1-\alpha_i^{nf}\right)\left(rvu_{i,w}^{nf} a_{i,w}^{nf}+rvu_{i,pe}^{nf} a_{i,pe}^f+rvu_{i,mp}^{nf} a_{i,mp}^f\right)\end{bmatrix}a_{CF}CF \tag{0.35}$$

From the perspective of the governing entity, a budgetary restraint is imposed such that a financial savings, *S*, is the goal for a time period. If the utilization of service code *i* is n_i , then aggressively,

$$\sum_{i\in P} n_i R_i \alpha_i = S(\alpha) \geq S \qquad (0.36)$$

If $\left|S(\alpha)-S\right| = \Delta(\alpha)$ represents governmental agency latitude in rate reduction negotiating, then a dual problem arises between the sustainability of the healthcare entity, (i.e., maximizing a facility's *gpm*) and maximizing $\Delta(\alpha)$ subject to sustaining patient accessibility to services. Patient accessibility does not equate to each n_i being sustained based on a static population because a results and evidence-based framework is normally utilized as better indicators of successful accessibility. Hence, the number of successful episodes and shorter lengths are better indicators. For chronic episodes, maintainability of goals met is a truer indicator of successful accessibility. In patient histories, the increase of goal status and maintainability uniformly done is the gold standard for quality accessibility and true capacity of care. Finally, delay times, $\Delta_d = soc_d - r_d > 0$, from the referral date to the start of care date (SOC) is a gross indicator of general patient accessibility. Accessed from historical claims data, a marked increase in the average difference, $\bar{\delta}(\alpha)$ between evaluation/treatment delay times before and after rate reduction implementation, $\delta(\alpha) = \Delta_d(\alpha) - \Delta_d(0)$ would indicate an accessibility pathology. $\bar{\delta}(\alpha)$ is dependent on patient SOC queue times for a coalition of facilities. If the cost adjustment functions, $d(\alpha) = \left[d(\alpha_i)\right]_{i\in P}$ dictate that a facility's staff of qualified providers is reduced by $N_{d(\alpha)}$ in order to be sustainable, both the patient wait queue time and throughput rate will accordingly increase. Consequently, the patient *capacity* of the healthcare entity will decrease as per those increased parameter levels.

We assume that patient care episodes follow a queuing system, $A/S/n$, where *A* is the arrival process, *S* is the service process and *n* is the number of servers, (i.e., number of providers in practice) (Cote and Stein, 2007) . The rate-reduced queuing system will be indicated by $A/S/\left(n - N_{d(\alpha)}\right)$. Service times are episodic, not encounter or visit-based, (i.e., the episode defines the total service time for a diagnosis). Patient care episodes are multi-channel multi-phase processes since different treatments and providers (servers and service types) may be involved. One would then approximate the difference between the average number of patients in service in a $A/S/n$ process and a $A/S/\left(n - N_{d(\alpha)}\right)$ process. In healthcare service queues these processes are approximated using *A* and *S* as Markov processes, *M* (Singh, 2006). Denote the difference in average number of patients between the two service processes by $\Delta\bar{L}_n(\alpha)$. Denote the difference in average wait times of the two processes by $\Delta W_n^q(\alpha)$. $\Delta W_n^q(\alpha)$ can then be estimated from *A* and *S*. $\Delta W_n^q(\alpha)$ can also be estimated by using the historic delay times $\bar{\delta}(\alpha_i)$ from above. $\Delta W_n^q(\alpha)$ is then an indicator (estimator) of access to care decline in terms of increased average wait times given the rate reduction percentage α . Delayed treatments equate to delayed access to care. If there is a large enough delay, treatment may never happen. We call this *tipping point* delay the *critical wait time* and denote it by W_c . W_c is a stochastic process as there will never be a deterministic critical wait time for all situations. Nonetheless, if $\Delta W_n^q(\alpha) > W_c$ *a. e.*, one may surmise that a number of treatments, nt_c , otherwise executable, will not be rendered on average during a typical work period.

Measurements of access to care robustness are usually defined by payors as network provider capacity or vacancy rates, in terms of an inadequate number of providers per a unit number of patients in the regional network. As the number of servers or providers *n*, in a network decrease ($n \to 0$) because of non-participation or non-availability, $\Delta W_n^q(\alpha) \to \infty$. Given that access to care is measured in provider per capita percentages, a certain percentage reduction, say μ %, is considered critical in government agency policy, that is, a decrease of μ % in per capita provider network coverage. Hence, using this standard, if $N_{d(\alpha)} \geq \mu n$, the network has reached an access to care crisis. However, one may reach an access to care shortage long before the μ % decrease in providers happens if $\Delta W_n^q(\alpha) > W_c$ *a. e.* for a smaller reduction α. We therefore suggest that the condition $\Delta W_n^q(\alpha) > W_c$ *a. e.* is a more robust indicator of an access to care crisis than the condition $N_{d(\alpha)} \geq \mu n$. It thus remains to obtain a robust estimate of W_c given historical data, if possible. This data can be gleaned from cancelation rates and reasons for cancelation from patient records. One may then more robustly estimate μ by:

$$\mu_c^n(\alpha) = \arg\min_{\mu}\left(N_{d(\alpha)} \geq \mu n;\ \Delta W_n^q(\alpha) > W_c\ \ a.\,e.\right) \tag{0.37}$$

In addition to the concept of decreased providers (servers) leading to decreased patient flows and hence to a decreased access to care, other healthcare resources, such as equipment, rooms, and other ancillary staff may be reduced with decreased reimbursements, as the healthcare business adapts to decreased revenues from reimbursement reductions. *Clinical pathways* dictate patient flows in generalized clinical setting. They require mathematical models to objectively describe their dynamics. To this end, process algebras (PA) pose as generalized mathematical models to determine the dynamics of *clinical pathways* in a healthcare services business. Process algebras are abstract algebraic models for state process flows. When stochastic or uncertainty mechanisms are involved in patient flow components, these process algebras can be equipped with stochastic processes within the algebraic semantics of the PA, creating stochastic process algebras. Recently, a specialized stochastic process algebra was proposed for assessing robust performance analysis in healthcare settings, the *clinical pathways performance evaluation process algebra* (CPP) (Yang, et al, 2012). In this model, patients are treated as flow particles in a process that travels from one service location to another with state information. The patient particle flow can be concurrent to servers (multiple ancillary, specialists, and primary providers) and resources. The CPP is a concurrent state process algebra that considers servers and resources and their interactions in a stochastic manner.

For our purposes, we are interested in utilizing the CPP to model and predict differences in the dynamics of a healthcare services setting when servers and resources change based on revenue drops from reimbursement reductions. We want to estimate the reduction in patient particle flow as a stochastic function of resource and server (provider) reductions. This flow decrease will then cause an access to care reduction as a function of the stochastics of resources and servers.

We refer to the Appendix for the definition of a clinical pathway and its components and semantics. Let $CP = \langle S, R, A, C, \digamma_C \rangle$ be the space of clinical pathways in a healthcare services entity. The space of all resources R, contains all resources such as providers and staff, equipment, rooms, computational devices, etc. needed to address the healthcare service concerns of the patient in the healthcare services entity. The space of states S, contains all possible patient location/stage states in the healthcare services entity. The space of actions A, contains all possible activities (as pairs of action types and their corresponding rates of activation) that are actualized on patients. The space of constraints C, are constraint conditions on all components from S and R. Finally, the space of functions F_C , define action rates as functions of elements in C.

The difference in system process (patient) throughput in a healthcare services entity dictates the reduction in services from a rate reduction η . Let $T\left(R(\eta)\right)=\alpha_d U_{R_N(\eta)}$ depict the measurable throughput with a reduced resource space $R(\eta)$ as a result of a rate reduction η . Define the difference in throughputs as $\Delta_\eta^T = T(R) - T\left(R(\eta)\right) = \alpha_d\left(U_{R_N} - U_{R_N(\eta)}\right)$, where α_d is the discharge action taken upon a patient in the system and $U_{R_N(\eta)}$ is the expected utilization of resources $R_N(\eta)$ after rate reduction η . Δ_η^T is then estimable and measures the reduction in patient throughput and hence the reduced or lack of access to care for those patients not able to be serviced by the new capacity of the healthcare service entity. See the appendix for details on the development of estimators for utilization and throughput in a healthcare services entity modeled as a performance evaluation process algebra. One may then utilize the condition from (0.37) using Δ_η^T instead as:

$$\mu_c^n(\eta) = \arg\min_\mu\left(N_{d(\eta)} \geq \mu n;\ \Delta_\eta^T \geq \omega\left(\alpha_d U_{R_N}\right)\right) \tag{0.38}$$

where ω is a predetermined threshold ratio that defines a critical throughput reduction (i.e., $\omega = 0.5$ defines a 50% reduction in patient throughput because of reduced resources).

Market Dynamics

When a rate reduction fee schedule is proposed by a government entity, review periods are established in which the effected healthcare sector may negotiate by issuing either a counter-proposal for a fee schedule and/or justification for retaining higher rate schedules based on economic sustainability directly influencing general patient accessibility to care. Let $\alpha = \left(\alpha_i\right)_{i\in P}$ denote a proposed rate reduction schedule. By relating any of the firm's KPIs, such as *gpm*, to sustained patient accessibility to care, a counter-proposal, $\alpha^c = \left(\alpha_i^c\right)_{i\in P}$ may be approximated using the following mathematical (linear initially) programming problem:

$$\begin{gathered}\max_{\alpha^c} gpm_c^*(O_j,U)\left(\alpha^c\right),\\ \text{subject to } \frac{\sum_{p\in U}\sum_{i\in I_{O,p}}\left(1+\alpha_i^c-\alpha_i\right)R_i}{\sum_{p\in U}\sum_{i\in I_{O,p}}\left(1+\alpha_i^c-\alpha_i\right)} < D\\ \Delta\left(\alpha^c\right)\geq 0,\ \text{where } \alpha^c = \left(\alpha_i^c\right)_{i\in I_{O,p}}\\ 0\leq \alpha_i, \alpha_i^c \leq 1, i\in I_{O,p}\end{gathered} \tag{0.39}$$

where D is a threshold for a reimbursement amount-weighted average rate reduction difference. The constraint may also be weighted by the relative frequencies of service for each CPT code multiplied by the reimbursement amount, $\frac{\sum_{p \in U_O} \sum_{i \in I_{O,p}} \left(1+\alpha_i^c - \alpha_i\right) f_{p,i} R_i}{\sum_{p \in U_O} \sum_{i \in I_{O,p}} \left(1+\alpha_i^c - \alpha_i\right)} < D$. The threshold D may then be used as a negotiating baseline. In terms of the effective rate reduction expressed in (0.3), the optimized rate reduction schedule $\alpha^c = \left(\alpha_i^c\right)_{i \in I_O}$, translates to the new optimized effective rate reduction schedule:

$$\bar{\alpha}_w^c = \frac{\sum_{p \in U_O} \sum_{i \in I_{O,p}} f_i R_i \alpha_i^c}{\sum_{p \in U_O} \sum_{i \in I_{O,p}} f_i R_i} \tag{0.40}$$

In the expression for $gpm^*(O_j, U)(\alpha)$, the cost adjustment functions, $d_i\left(\alpha_i\right)$ are also constrained so that patient accessibility is not reduced. Recall that $n_p q_{p,i}$ is an estimate for the number of insurees in the facility's service area from the payor p network. The number of serviced insurees in the service area is estimated as $\sum_{O \in \Omega_O} \sum_{p \in U_O} \sum_{i \in I_{O,p}} n_p q_{p,i} ms_{p,i}^O$, where $ms_{p,i}^O$ is the market share estimate of payor p insurees for service i, for facility O. The sum is taken over all competitors, O in the facility's service area Ω_O and all serviced CPT codes i. Now assume that after the rate reduction schedule α, has been implemented, all competitors, $O_k \in \Omega_O$ have adjusted their respective costs according to their respective cost adjustment functions, $d_i^{O_k}\left(\alpha_i\right)$. In a rational game, each competitor $O_k \in \Omega_O$, will maximize their respective $gpm^*(O_k, U_{O_k})\left(\alpha\right)$. However, what has not been exposed is a competitor's reserve resources, (i.e., flexible holdings). Under this scenario, each competitor holds separate flexible holdings, H_{O_k}. Then a game strategy space for O_k will include the possible actions: (1) folding the facility, (2) forming a coalition with other facilities, (3) selling the facility, and (4) buying and consolidating with a separate facility (facilities). These actions will in turn change market shares in the service area. Tobin's Q measure can then be used as an approximate barometer for directing the action to be taken. H_O can be included in the calculation of Q_T as an added component. Also, Q_T will change as a rate reduction schedule α, is introduced. Let $Q_T(O,U,\alpha)$ denote the Tobin measure after a rate reduction, α and $Q_T(O,U,0)$ be the baseline Tobin measure respectively under payor portfolio U. The following action space is proposed under such scenarios:

fold O	if $Q_T(O,U,\alpha) << Q_T(O,U,0)$ and $Q_T(W,U,\alpha) < Q_T(W,U,0)$, $\forall W \subset \Omega_O$, subcoalitions, $\ni W \cap O = \varnothing$
form coalition	if $Q_T(O,U,\alpha) << Q_T(O,U,0)$ and $Q_T(W,U,\alpha) \geq Q_T(W,U,0)$, for some subcoalition, $W \subset \Omega_o$ $\ni W \cap O = \varnothing$
sell O	if $Q_T(O,U,\alpha) << Q_T(O,U,0)$ and $Q_T(O,U,\alpha) \geq Q_T(O \cup W,U,\alpha)$, for some subcoalition, $W \subset \Omega_o$, $\ni W \cap O = \varnothing$
buy W	if $Q_T(O,U,\alpha) < Q_T(O,U,0)$ and $Q_T(O,U,\alpha) < Q_T(O \cup W,U,\alpha)$, for some subcoalition, $W \subset \Omega_o$, $\ni W \cap O = \varnothing$

(0.41)

The market price, $mp\left(O_j\right)$ of a facility O_j, used in defining *Q*, increases as market share $ms\left(U_O\right)$, increases relative to its competitors. Again, see the Appendix for a definition of market share of a healthcare firm in a competitive space. The asset price, a_{U_O} in *Q* may or may not increase as market share increases because the appropriate cost adjustment may not be effective in increasing $gpm^*(O_j, U)(\alpha)$. Let $a^O_{U_O,\alpha}$ denote the asset price of facility *O* under payor portfolio U_O and after rate reduction schedule α has been implemented. One may express $a^O_{U_O,\alpha}$ as a function of $g^O_{U_O,\alpha} = gpm^*(O,U)(\alpha)$, the *gpm* of facility *O* after rate reduction schedule α has been implemented and under payo*r portfolio*, U_O, as $a^O_{U_O,\alpha} = h\left(g^O_{U_O,\alpha}\right)$. An adjusted market price $mp\left(O,\alpha\right)$, to $mp\left(O\right)$ is more intangible, but also more realistic, as it is a function of the status of the industry, facility dynamics, and the general economy (world, national, state, and local leading economic indicators) after rate reduction schedules have been implemented. Nonetheless, economic indicators directly influence all costs and revenues and as such, influence $g^O_{U_O,\alpha}$ as well. Hence, one may assume a functional relationship, $mp\left(O,\alpha\right) = u\left(g^O_{U_O,\alpha}\right)$. Finally,

$$Q_T(O,U_O,\alpha) = \frac{mp\left(O,\alpha\right)}{a^O_{U_O,\alpha}} = \frac{u\left(g^O_{U_O,\alpha}\right)}{h\left(g^O_{U_O,\alpha}\right)} = H\left(g^O_{U_O,\alpha}\right) \qquad (0.42)$$

where $H \underset{def}{\equiv} \frac{u}{h}$ is a rational function (assuming the healthcare firm has nonzero tangible assets). The parametrization, $\Pi_O = \left(\alpha, U_O, d, H_O\right)$ will then dictate the facility's sustainability. One would then negotiate rate reductions based on the solution to (0.39) with $g^O_{U_O,\alpha} = gpm^*_c(O,U_O)(\alpha)$ replaced by $Q_T(O,U_O,\alpha)$.

Risk Assessment

We conclude with a strategy for risk assessment calculations. In calculating sums involved in *gpms* and effective rate reductions, one can introduce a *perception factor* or *influence* for each service code, *i*, that would reflect perceptions of added emphasis or deemphasis. These perceptions are reflections of where on a risk management spectrum, (i.e., risk-aversion to risk-aggression), firms are advocating. Let these perception factors be given by $\phi_{U_O} = \left(\phi_{i,p}\right)_{p\in U_O, i\in I_{O,p}}$ for payor *p* reimbursement schedules. Then, for the optimization problem (0.39),

$$gpm^*_c(O,U_O)\left(\alpha\right)\left(\phi\right) \approx 1 - \frac{\sum_{p\in U_O}\sum_{i\in I_{O,p}} \phi_{i,p}\hat{c}^*_{p,i}(\alpha_i)R_i(\alpha_i)}{\sum_{p\in U_O}\sum_{i\in I_{O,p}} \phi_{i,p}\hat{r}_{p,i}(\alpha_i)R_i(\alpha_i)} \qquad (0.43)$$

The carveout effective rate reduction can then be recalculated as:

$$\bar{\alpha}_w^c(O,U_O)(\phi)=\frac{\sum_{p\in U_O}\sum_{i\in I_{o,p}}\phi_{i,p}f_{i,p}R_{i,p}\alpha_i^c}{\sum_{p\in U_O}\sum_{i\in I_{O,p}}\phi_{i,p}f_{i,p}R_{i,p}} \tag{0.44}$$

The loss (without cost adjustments) as a result of a rate reduction schedule $\alpha=(\alpha_i)_{i\in P}$, is:

$$\begin{aligned} L(O,U_O,\alpha) &= \sum_{p\in U_O}\left[\sum_{i\in I_{O,p}} r_{p,i}(\alpha_i)R_i(\alpha_i)-\sum_{i\in I_{O,p}} r_{p,i}R_i\right] \\ &= \sum_{p\in U_O}\left[\sum_{i\in I_{O,p}}\left(r_{p,i}(\alpha_i)-r_{p,i}\right)R_i-\sum_{i\in I_{O,p}}\alpha_i r_{p,i}(\alpha_i)\right] \end{aligned} \tag{0.45}$$

Now define the probability density describing the rate reduction choice made by government entities, $p(\alpha)=\text{ probability } \alpha \text{ choosen}$. The risk involved in negotiating a rate reduction adjustment, as an average expected risk is:

$$\begin{aligned} R &= E_\alpha\left[L(O,U_O,\alpha)\right] \\ &= \int_{H_N} L(O,U_O,\alpha)p(\alpha)d\alpha \\ &= \sum_{p\in U_O}\int_{H_N}\left(\sum_{i\in I_{O,p}}\left(r_{p,i}(\alpha_i)-r_{p,i}\right)R_i-\sum_{i\in I_{O,p}}\alpha_i r_{p,i}(\alpha_i)\right)p(\alpha)d\alpha \end{aligned} \tag{0.46}$$

The integral is taken over the $N=|P|$ dimensional hypercube $H_N=\left\{(\alpha_i)_{i\in I_{O,p}}:0\le\alpha_i\le 1\right\}$. Beyond a certain range of rate reductions, the loss becomes disproportionate since a certain business closing would happen. However, since a carve-out can be proposed, this *danger zone* region in H_N resembles a hyper-parallelepiped (different lengths for sides) instead of a hypercube. Let $D_N=\left\{(\alpha_i)_{i\in I_{O,p}}:\varphi_i\le\alpha_i\le 1\right\}$ denote this danger zone region where $\varphi=(\varphi_i)_{i\in I_{O,p}}$ is the vector representing the service code component-wise thresholds for danger. In D_N, the loss is significantly elevated, even catastrophic. So, $L(O,U_O,\alpha)$ should be proportionately elevated in D_N. To this effect, let $B(\alpha)=\left(B(\alpha_i)\right)_{i\in I_{O,p}}$, where $B(\alpha_i)\ge 1$, such that $B_i(\alpha_i)\to\infty$ as $\alpha_i\to{}^{-}1$ be *bump factors* when the loss approaches or has penetrated the danger zone. Then (0.46) can be rewritten, considering the elevated losses, as:

$$\begin{aligned} R = &\sum_{p\in U_O}\int_{H_N\setminus D_N}\left(\sum_{i\in I_{O,p}}\left(r_{p,i}(\alpha_i)-r_{p,i}\right)R_i-\sum_{i\in I_{O,p}}\alpha_i r_{p,i}(\alpha_i)\right)p(\alpha)d\alpha + \\ &\sum_{p\in U_O}\int_{D_N}B(\alpha_i)\left(\sum_{i\in I_{O,p}}\left(r_{p,i}(\alpha_i)-r_{p,i}\right)R_i-\sum_{i\in I_{O,p}}\alpha_i r_{p,i}(\alpha_i)\right)p(\alpha)d\alpha \end{aligned} \tag{0.47}$$

The bump factors $B\left(\alpha\right)$ can be implemented through internal healthcare firm strategy changes and rate re-negotiations with payors. Re-negotiation is leveraged by community and political support, variety and quality of services offered, credentials, number of competitors (provider saturation levels), and the size of the payor service subpopulation within the healthcare firm's patient rolls. In terms of patient accessibility, the loss in revenues (given fixed costs) can be equated with reduced service frequencies as in the following:

$$\left(r_{p,i}(\alpha_i)-r_{p,i}\right)R_i-\alpha_i r_{p,i}(\alpha_i)=f_{p,i}s_{p,i}R_i \quad (0.48)$$

or directly as decreased service frequencies,

$$\begin{aligned} f_{p,i} &= \frac{\left(r_{p,i}(\alpha_i)-r_{p,i}\right)R_i-\alpha_i r_{p,i}(\alpha_i)}{s_{p,i}R_i} \\ &= \left[f_{p,i}\mu_{i,p}-R_i\right]-\alpha_i\left[\frac{f_{p,i}\mu_{i,p}(\alpha_i)}{R_i}\right] \end{aligned} \quad (0.49)$$

Substituting and simplifying, this becomes,

$$\begin{aligned} f_{p,i} &= \frac{R_i}{\mu_{i,p}(\alpha_i)\left(1+\frac{\alpha_i}{R_i}\right)-1} \\ &= \frac{R_i}{\frac{\tau_{i,p}(\alpha_i)}{(1-\alpha_i)}\left(1+\frac{\alpha_i}{R_i}\right)-1} \end{aligned} \quad (0.50)$$

recalling that $\tau_{i,p}(\alpha_i)$ is the adjusted payor rate reduction developed from the baseline rate reduction, α_i.

Conclusions

The dynamics of Medicaid reimbursement rates rival those of Medicare. Medicare has always been the harbinger of provider reimbursements in all insurance programs using the RBRMS standard. However, Medicaid reflects the regional changes of healthcare more readily. In this paper we have laid out several technical developments in correlating and explaining how reimbursement rates correlate and affect healthcare business financial stability, viability, and robustness. These effects then dictate access to care measurements and local market dynamics for the healthcare provider. We have shown how one stimulus, a Medicaid rate reduction percentage, can affect key performance indicators for a healthcare business which cascade to access to care and market changes. These microeconomic metrics may help to measure the true effects of policy changes in the Medicaid landscape and dictate the spectrum of dynamics in its implementation.

Appendix

Market Share Dynamics

Consider calculating an approximation to market share for a healthcare firm within its service area. Let O_j be the *j-th* firm within its service area, SA_{O_j}. Let n_{O_j} be the number of eligibles in SA_{O_j}. Define a divergence measure, $d^*\left(i,O_j\right)$ between the *i-th* eligible in SA_{O_j} and O_j as follows. This measure defines the marketable proximity of *i* to O_j. First, define the *stickiness* of *i* to O_j as a measure of retention or attraction, $s(i,O_j)$. This measure could include brand loyalty (preference to provider and methodologies used), comfortability, cost preference, and attraction to localization of special services. Next, define the physical navigable distance between *i* and O_j by $d(i,O_j)$. Finally, define the availability or accessibility of O_j to *i* by $a(i,O_j)$. This may include the accessibility by the simultaneous sub-capacity of the firm, the ability of the patient to transport to/from the healthcare firm, and payor mutual participation. Define $d^*\left(i,O_j\right)$ by these measures as follows:

$$d^*\left(i,O_j\right)=\frac{d(i,O_j)}{s(i,O_j)+a(i,O_j)} \tag{0.51}$$

Now define the ε*-potential* market share of *i* to O_j by:

$$ms(i,O_j,\varepsilon)=\begin{cases}\dfrac{1}{n\left(i,O_j,\varepsilon\right)}, & \text{if } d^*(i,O_j)\le \min\limits_{k\in\Omega_{O_j}} d^*(i,O_k)+\varepsilon\\ 0, & \text{otherwise}\end{cases} \tag{0.52}$$

where $n\left(i,O_j,\varepsilon\right)=\left|\left\{O_m\in\Omega_j \mid d^*(i,O_m)\le \min\limits_{k\in\Omega_{O_j}} d^*(i,O_k)+\varepsilon\right\}\right|$ and Ω_{O_j} is the competitor space around facility O_j. Finally, define the service area ε*-potential* market share of O_j:

$$ms(O_j,\varepsilon)=RND\left(n_{O_j}\sum_{k\in\Omega_{O_j}} ms(k,O_j,\varepsilon)\right) \tag{0.53}$$

where *RND* is the integer rounding function. As $\varepsilon\to 0$, $ms(O_j,\varepsilon)$ approaches a crisp definition of market share, when it exists:

$$ms\left(O_j\right)=\lim_{\varepsilon\to 0} ms(O_j,\varepsilon) \tag{0.54}$$

Practically, there is usually a minimum ε, $\varepsilon^*>0$, where $ms\left(O_j,\varepsilon^*\right)=ms\left(O_j,\varepsilon\right),\forall\varepsilon<\varepsilon^*$ (i.e., infinitesimal fluctuations in the market share metrics used do not further influence market share).

Clinical Pathway Performance Evaluation Process Algebras (CPP)

Let $CP = \langle S, R, A, C, \mathsf{F}_C \rangle$ be the space of *clinical pathways* (*CP*s). *CP*s are mathematical objects that represent stochastic process models for patient state particle flows (throughput) in concurrent service conduits such as healthcare service entities involving resources. We follow (Yang, et al, 2012) and give some clarity and generalizations in the sequence below. Components of a clinical pathway are:

(1) S , a finite set of states, $s \in S$ indicating the state locations where patients are being treated;

(2) R , a finite set of resources (providers, staff, rooms, utilities, equipment, etc.), $r \in R$ required for patient treatments;

(3) A , a finite set of activities (actions), $a \in A$ that are performed on patients that change resources for patient states and patient states and where $a = (\alpha, \lambda_\alpha)$, α is an action type from the space of all possible action types AT , and $\lambda_\alpha > 0$ is a rate of action execution defined as $\lambda_\alpha = \frac{1}{d_\alpha}, d_\alpha = E(\mu_\alpha)$ where μ_α is execution duration time distributed as $\mu_\alpha \sim \exp(\beta_\alpha)$ for some parameter β_α (i.e., λ_α is negative exponentially distributed);

(4) $C \cup \{NULL\}$, is a set of constraints, $c \in C$, put upon resources, states, and actions, where $NULL$ is the null constraint (passthrough constraint); and

(5) F_C , a set of functions, $f \in \mathsf{F}_C; f : C^n \to R^+ \cup \{0\}$, that determine action rates $\lambda = f(c_1, c_2, ..., c_n)$ where $c_i \in C,\ 1 \le i \le n$, for some $n \in \mathbb{N}^+$.

For purposes of defining the language semantic operators of the process algebra of CPP, let $p, q, t, u, v, w, x \in R(or\ S)$ (resources or states), and $a \in A$ (action) where $a = (\alpha, \lambda_\alpha)$. The (functional) state of the resource *p* given by the process (0.55) below after the action a and those of the action subspace $L \subseteq A$ with respect to *w* and *x*, have been executed with respect to the (functional) states of the resource (or state) q, v, w, x , is given by the process algebra notational semantic (definition):

$$p \overset{\text{def}}{\equiv} (\alpha_a, \lambda_a).q \mid u + v \mid w \underset{L}{\rhd\!\lhd} x \qquad (0.55)$$

where the *sequential operator* $(.)$, defines the sequential order (prefixing) of processing of the left and operands; the *choice operator* $(+)$, defines a competition relationship (sum of possibilities) between the left and right operands; the *cooperation operator* $\left(\underset{L}{\rhd\!\lhd}\right)$, defines an interaction between the left and right operands over the shared action subset $L \subseteq A$, which is usually parallel compositional (simultaneous processing at the slowest rate of the two being the overall terminal rate); and the abstraction operator $(|)$, hides or makes private the operand to the right of the operator with respect to the operand to the left of the operator (i.e., the abstraction of the right operand makes it *invisible* to outside processes and components when interacting with the left operand). An additional *combinator* process operator $(||)$, is defined as a form of null cooperation when two components behave completely independently (no cooperation) or equivalently $\left(\underset{\varnothing}{\triangleright\triangleleft}\right)$. The definition assignment $\left(\overset{\text{def}}{\equiv}\right)$ assigns names to a component that matches patterns of behavior associated with the component statement on the right side. Examples of component statements and their meaning in process flows include:

(1) $p \overset{\text{def}}{\equiv} (\alpha, \lambda).q$; component *p* becomes *q* with the completion of activity (α, λ).

(2) $p + q$; behaves either as *p* or *q*, enabling all activities of *p* and *q* with the first completed activity determining the final component as *p* or *q*.

(3) $p \underset{L}{\triangleright\triangleleft} q$ defines a process that forms a compositional (cooperating) function of *p* and *q* when the shared activities in *L* are activated. A simple example of cooperating composition is when the two components *p* and *q* sharing data from each other (synchronizing) and process in parallel with terminating coming after the slowest component terminates.

(4) Introducing stochasticity, let $\rho_i \sim g_i\left(\overline{\theta}_i\right) i = 1, 2, \ldots m$, be *m* distributed random variables. Then $p_j \overset{\text{def}}{\equiv} \sum_{i=1}^{m} (\alpha, \rho_i \lambda_\alpha).p_i,\ j \notin \{1, 2, \ldots, m\}$ is a component (patient) that transfers from a state p_j at location *j* to a state $p_i, i = 1, 2, \ldots, m$ at location *i* with probability ρ_i respectively, after completing an activity with action type α.

(5) Define a resource functional (capacity) state by $R_j\left(\kappa_{ij}\right), j = 1, 2, \ldots, l; i = 1, 2, \ldots, m_j$, needed respectively for the flow particles (patients) state, p_j at locations $i = 1, 2, \ldots, m_j$ where $0 \le \kappa_{ij} \le 1$ are the m_j ordered spectral capacity states, (i.e., $0 = \kappa_{1j} < \kappa_{i+1,j} < \ldots < \kappa_{m_j j} = 1$). Then $R_j\left(\kappa_{ij}\right) \overset{\text{def}}{\equiv} \left(\alpha_{il}, \lambda_{\alpha_{il}}\right).R_j\left(\kappa_{lj}\right), i, l = 1, 2, \ldots, m$, represents the resource capacities where for location *j*, the resources $R_j\left(\kappa_{ij}\right)$ needed are at κ_{ij} capacity respectively when completing activities of action types $\left\{\alpha_{il} : i, l = 1, 2, \ldots, m_j\right\} \subseteq A$ respectively. The corresponding ordered patient state transitions can then be expressed as $p_{j-i} \overset{\text{def}}{\equiv} \left(\alpha_{il}, \lambda_{\alpha_{il}}\right).p_{j+l-i+1}$. This state transition represents an over generalization for most healthcare services entities since usually only two resource capacities are relevant - idleness and total occupancy between every consecutive patient location/stage state.

(6) To describe the (whole) system state (overall clinical pathway), one must describe the relationships between the patient flow states and resource states in terms of a cooperation

operator (i.e. patient treatments need a certain level of resources, while resources are at a certain capacity level as treatment requirements for those patients evolve). We define this cooperation as:

$$p_{j+i} \overset{\text{def}}{\equiv} p_j \underset{\{\alpha_{p_{j+i}},\alpha_{R_{j+i}}\}}{\triangleright\triangleleft} R_{j+i}\left(\kappa_l\right) \tag{0.56}$$

where $\left\{\alpha_{p_{j+i}},\alpha_{R_{j+i}}\right\}$ are the respective (treatment) action types who upon execution, p_j and R_{j+i} at capacity state κ_l, synchronize to produce p_{j+i}. This patient location/stage state transition can be interpreted as follows: patient at location/stage *j* is transitioned to location/stage *j+i* via a treatment action type α_{p_j} when treatment resources at location/stage *j+i* are at capacity κ_l via treatment action type $\alpha_{R_{j+i}}$ at that location/stage.

In a healthcare services entity, there are multiple patient flows and resources. One must then define the location/stage state of a patient as a function of the number of patients in the queued location/stage state, N_p and the capacity state of a resource as a function of the number of resources, N_R in that capacity state necessary for that patient location/stage state using the parallel combinator:

$$\begin{aligned} p_j\left(N_p\right) &\overset{\text{def}}{\equiv} \underbrace{p_j \| \ldots \| p_j}_{N_p} \\ R_j\left(N_r\right) &\overset{\text{def}}{\equiv} \underbrace{R_j \| \ldots \| R_j}_{N_r} \end{aligned} \tag{0.57}$$

Here, the j^{th} patient location/stage state is given by the parallelization of N_p individual j^{th} location/stage states and the j^{th} resource capacity state is given by the parallelization of the N_R j^{th} capacity state resources for those parallel j^{th} patient location/stage states.

To create greater patient flow efficiencies in a healthcare services entity, one must, at times, pre-queue patient location/stages (e.g., patients pre-queued in waiting rooms rather than waiting outside, patients pre-queued in treatment rooms or surgical-prep areas before providers appear, etc.). Together with the consideration of parallel patient care, we may express a new composited cooperation operator defining a new patient location/stage state for the system (clinical pathway) with pre-queued states:

$$\begin{aligned} p_{j+i}\left(N_p\right) \overset{\text{def}}{\equiv} p_j\left(N_p\right) &\underset{\{\alpha_{R_l}\}}{\triangleright\triangleleft} R_l\left(\kappa_a\right)\left(N_R\right) \\ &\underset{\{\alpha_{p_{j+i}},\alpha_{R_{j+i}}\}}{\triangleright\triangleleft} R_{j+i}\left(\kappa_b\right)\left(N_R\right) \end{aligned} \tag{0.58}$$

for some capacity indices *a* and *b*, and intermediate pre-queued patient location/stage state *l*, $i < l < i + j$. To accommodate the transition into pre-queued positions, the resource capacities must be sufficient (e.g., if N_{wait} depicts the capacity of the waiting area, then $N_{wait} \geq N_p$ and $\alpha_{R_l} \square \alpha_{R_{j+i}}$, quick transition to new location/stage when resources become available at that new location/stage, minimizing patient location/stage transition lag time).

The CPP model inherits Markov processes that determine the flow from one state to another. Markov processes are memoryless state transitions, (i.e., the time elapsed to the next transition is independent of the time elapsed to the current state from other past states). In the CPP model one defines these transition time rates in terms of the activity rates of actuation α_i. Let $S_i, S_j \in S$ be two distinct states. The state transition rate between S_i and S_j is given by:

$$q_{S_i,S_j} = \sum_{\{\alpha \in AT: S_i \to S_j\}} \alpha \tag{0.59}$$

where the activity type index set $\left\{\alpha \in AT : S_i \to S_j\right\}$ in the summation above is the set of all activity types that transition S_i to S_j. In continuous-time Markov transitions, one can use infinitesimal time increments δt to define infinitesimal transition probabilities about a stochastic process that represents when the system is in a state S_i. For example, define the stochastic process that defines when the system is in state S_i at time t by $X(t) = S_i$. Then $P\left(X(t+\delta t) = S_j \mid X(t) = S_i\right) = q_{S_i,S_j}\delta t + O(\delta t), i \neq j$ defines the probability of transitioning from S_i to S_j within a time interval δt in terms of the process (cumulative) transition rate and the time interval. One uses these continuous transition probabilities to define the evolution of a continuous Markov process in the form of a first-order differential equation:

$$\frac{\partial P(t)}{\partial t} = P(t)Q,\ P(0) = 1_N \tag{0.60}$$

where Q is the square generator matrix for the transition rates given by off-diagonal elements $q_{S_i,S_j}, i \neq j$ and diagonal elements $q_{S_i,S_i} = -\sum_{i \neq j} q_{S_i,S_j}$, $P = \left(p_{i,j}\right)_{i,j}$ is the square matrix of the transition probabilities $p_{i,j} = P\left(X(t) = S_j \mid X(0) = S_i\right)$, and 1_N is the square identity matrix of rank N, where N is the number of states in S.

The steady-state behavior (through the steady state distribution of these transition probabilities via (0.60), when it exists) of the system of clinical pathways displays the long-term equilibrium of the system and hence of the stochastic measurement of throughput and resource capacities of the healthcare services entities as $t \to \infty$. To this end, the total time that the system (patient) spends in a state S_j is given by:

$$\pi_j = \lim_{t \to T} P\left(X(t) = S_j \mid X(0) = S_0\right) \tag{0.61}$$

where T is the total time the patient stays in the system (time from entrance to discharge). To obtain an equilibrium, the Markov process must be irreducible (i.e., every state must be reachable from every other state). Hence, we artificially assign the state when a patient is discharged S_d, as transitional to the entrance state S_0 with positive probability (i.e., $q_{S_d,S_0} > 0$). We may also extend our state space to the “home” state S_h that represents the patient being at home. At that point, the patient may return to the healthcare services entity entry state with some probability, $q_{S_h,S_0} \geq 0$.

We defined the steady-state distribution of the system as $\left[\pi_1 \cdots \pi_N\right]$. We next define the square matrix $\Pi = \lim_{t \to T} P(t) = \begin{pmatrix} \pi_1 & \cdots & \pi_N \\ \vdots & \ddots & \vdots \\ \pi_1 & \cdots & \pi_N \end{pmatrix}$ in order to change (0.60) into a matrix system:

$$\Pi' = \Pi Q = 0_N$$

$$\left[\pi_1 \cdots \pi_N\right] \begin{bmatrix} q_{S_1,S_i} \\ q_{S_1,S_i} \\ \cdot \\ \cdot \\ \cdot \\ q_{S_N,S_i} \end{bmatrix} = 0, \; i = 1, \ldots, N \tag{0.62}$$

This can be simplified to the condition:

$$\underbrace{\pi_i \sum_{j \neq i} q_{S_i,S_j}}_{\text{outflux from } S_i} = \underbrace{\sum_{j \neq i} \pi_j q_{S_j,S_i}}_{\text{influx to } S_i} \qquad (0.63).$$

Using (0.59), one can solve (0.63) for the steady-state distribution $\left[\pi_1 \cdots \pi_N\right]$.

Resource utilization R_j, at location/stage j may then be calculated by measuring the capacity states of the resource at the various system states S_i, where that resource is at a particular capacity and then summing those steady-state distribution values for those states. An average utilization can then be calculated if multiple copies of that resource exist. For example, for states in the state subspace $S(j, \kappa_l) \subset S$ where $R_j(\kappa_l)$ holds and if there are N_{R_j} copies of the resource R_j, then

$$U_{R_j} = \frac{\sum_{i \in \{k : S_k \in S(j, \kappa_l)\}} \pi_i}{N_{R_j}} \tag{0.64}$$

defines an average utilization of R_j at capacity κ_l in the steady-state of the system. If the action type α_d depicts the discharge action on a patient, then

$$T(R) = \alpha_d U_{R_N} \tag{0.65}$$

where N is the last (discharged) patient location/stage defines throughput of the system as a function of the space of resources R.

When $N_p + N_R \gg 100$, *state-space explosion* (number of possible states becomes computationally unwieldy) occurs for calculating steady-state solutions in (0.63). Approximation methods for solving continuous fluid flows described by differential equation systems of the form (0.62) for stochastic process algebra models come into usage (Bradley, et al, 2008). Let $N_{P_i(t)}$ denote the expected number of components (patients and resources) at location/stage *i* at time *t*. Using fluid dynamics, a differential equation can be generated to represent the evolution of $N_{P_i(t)}$ as follows:

$$\frac{\partial N_{P_i(t)}}{\partial t} = \underbrace{\sum_{\left\{j: p_i \underset{\alpha}{\rightarrow} p_j\right\}} \alpha}_{\text{outflux from p}_i} + \underbrace{\sum_{\left\{j: p_j \underset{\beta}{\rightarrow} p_i\right\}} \beta}_{\text{influx to p}_i} \tag{0.66}$$

The differential system (0.66) can then be formalized using the stochastic process algebra components defining the clinical pathway system in the righthand side expression of (0.66), along with the initial conditions of the system. Solving for (0.66), one obtains the continuous time evolution of each system component (patients and resources). Once $P_i(t)$ is approximated by that solution, mean (expected) times to peak capacity values can be calculated such as $M_i = E_t\left[\underset{t\in[0,T]}{\arg\max}\, P_i(t)\right]$, the expected time when peak patient capacity is reached. The expected passage time of a patient to be discharged from the time of entrance can also be approximated by attaching a *process state probe* in each patient that will activate when an activity α, has been performed on the patient. It is defined by the following component process:

$$\begin{aligned} &\text{probe}_{\text{off}} \overset{\text{def}}{\equiv} (\alpha, \lambda_\alpha).\text{probe}_{\text{on}} \text{ (probe first turned on)} \\ &\text{probe}_{\text{on}} \overset{\text{def}}{\equiv} (\alpha, \lambda_\alpha).\text{probe}_{\text{on}} \text{(probe left on after first activation)} \end{aligned} \tag{0.67}$$

The components p_i are then replaced by the components $p_i \underset{\alpha}{\triangleright\triangleleft} \text{probe}_{\text{off}}$. Let $G(\alpha) = \min_{t\in[0,T]} t : p_i(t) \underset{\alpha}{\triangleright\triangleleft} \text{probe}_{\text{on}}$ define the random variable representing the minimum time that a patient component has the action α performed on it. Then

$$F(\alpha) = F\left(G(\alpha)\right) \approx \sum_{\nu \underset{\alpha}{\triangleright\triangleleft} \text{probe}_{\text{on}} \in \Phi(S,R)} 1 \tag{0.68}$$

is an approximation to the CDF of $G(\alpha)$ where the righthand side of (0.68) is the count of all patient components of the form $\nu \underset{\alpha}{\triangleright\triangleleft} \text{probe}_{\text{on}}$ where ν are components formable in the system process algebra with state space *S* and resource space *R*, denoted by $\Phi(S,R)$. One can then approximate $E_t\left[G(\alpha)\right]$, the expected value of the random variable $G(\alpha)$. If α_d is the discharge action type, then $E_t\left[G(\alpha_d)\right]$ approximates the mean passage time to the discharge of a patient.

Of interest to this study is approximating the difference in patient capacity in a healthcare services entity because of a reduction of resources (providers, equipment, etc.) culminating from a reduction in revenues from an effective reimbursement rate reduction η . To this end, let $N_R(\eta)$ denote the number of resources available for use at the healthcare services entity after the rate reduction η is implemented. Let $N_R = N_R(0)$ denote the pre-reduction number of resources. We then want to estimate the difference in the expected times to capacity $M_i = E_t\left[\arg\max_{t\in[0,T]} P_i(t)\right]$ for each system, in order to estimate the difference in patient capacity times resulting from the reimbursement rate reduction η . Let M_i^{η} be the expected minimum time to patient capacity when rates have been reduced by η . Because of the subsequent reduction in resources to $N_R(\eta)$ from N_R , the resource capacities states $R_i(\kappa_l)$ will be met sooner, (i.e., $E_t^{N_R(\eta)}(R_i(\kappa_l)) \le E_t^{N_R}(R_i(\kappa_l))$ where $E_t^{N_R(\eta)}$ is the expected time to capacity κ_l , when resources have been reduced by a rate reduction of η . The difference, denoted by $\Delta E_\eta = E_t^{N_R}(R_i(\kappa_l)) - E_t^{N_R(\eta)}(R_i(\kappa_l))$ may then be estimated from the steady-state distribution. ΔE_η is then an estimate of the time dilation to capacity κ_l from a rate reduction η .

The difference in system throughput may also dictate the reduction in service from a rate reduction. Let $T(R(\eta)) = \alpha_d U_{R_N(\eta)}$ depict the measurable throughput with reduced resource space $R(\eta)$. Define the difference in throughputs as $\Delta_\eta^T = T(R) - T(R(\eta)) = \alpha_d\left(U_{R_N} - U_{R_N(\eta)}\right)$. Δ_η^T then measures the reduction in patient throughput and hence the reduced or lack of access to care for those patients not able to be serviced by the new capacity of the healthcare service entity.

References

Bradley, J. T., Hayden, R., Knottenbelt, W. J., & Suto, T. (2008). Extracting response times from fluid analysis of performance models. *Performance Evaluation: Metrics, models, and benchmarks 2008*.

ClearPoint Strategy (2017). Healthcare KPI library: 108 KPIs & scorecard measures. Retrieved from www.clearpointstrategy.com.

Cote, M. J. & Stein, W. E. (2007). A stochastic model for a visit to the doctor's office. *Mathematical and Computer Modelling, 45 (3-4), 309-323*.

Gapenski, L. C., & Reiter, K. L. (2016). *Healthcare finance: An introduction to accounting and financial management* (6th Ed). Chicago: Health Administration Press

Kaplan, R.S. & Norton, D.P. (1992). The balanced scorecard measures that drive performance. *Harvard Business Review*, January-February, 71-79.

MACPAC (2015). *An update on the Medicaid primary care payment increase*.

Nitzan, J. & Bichler, S. (2009). *Capital as power: A study of order and creord*er. London and New York: Routledge Press.

RAND Corporation (2017). *Examining the implementation of the Medicaid primary care payment increase*.

Singh, V. (2006). *Use of queuing models in healthcare*. University of Arkansas for Medical Sciences. Retrieved from http://works.bepress.com/cgi/viewcontent.cgi?article=1003&context=vikas_singh.

Sveiby, K.-E. (2010). *Methods for measuring intangible assets*. Retrieved from http://www.sveiby.com/articles/IntangibleMethods.htm.

Tobin, J. & Brainard, W. C. (1977). Asset markets and the cost of capital. In *Economic Progress, Private Values and Public Policy*. The Hague, Netherlands: North Holland.

Yang, X., Han, R., Guo, Y., Bradley, J., Cox, B., Dickinson, R., & Kitney, R. (2012). Modeling and performance analysis of clinical pathways using stochastic process algebra PEPA. *Workshop on Clinical bioinformatics, October 2012*.