NASMHPD RESEARCH INSTITUTE, INC.

Planning and Budgeting Evidence-Based-Practices for Mental Health and Substance Abuse Disorders

Anthony Broskowski, Ph.D.
Pareto Solutions
Palm Beach Gardens, Florida

Christina Thompson, Ph.D.
Heart of the Matter Consulting, Inc.
Ann Arbor, Michigan

Bill Barton
Pareto Solutions
Ithaca, New York

September 2004

INTRODUCTION

This document is intended to provide technical assistance to members of the National Association of Mental Health Program Directors (NASMHPD) on planning for and budgeting evidence-based programs (EBPs) for mental health and substance abuse treatment. Most EBPs do not provide clear cost estimates for replicating programs, or if they do, the costs are usually variable due to regional differences in salaries and other start-up and on-going operational expenses. This article provides a brief definition of EBP and description of the EBP known as Supported Employment. After reviewing some of the limits and errors associated with commonly used planning and budgeting methods, a method called stochastic (probabilistic) modeling is introduced, along with a discussion of how such models are designed and constructed. Following this discussion a simplified model, for the EBP of Supported Employment (SE) is introduced and used to illustrate the basic features of stochastic modeling. The stochastic model was designed and built partly based on information from a single article published by Latimer (5) that reported actual utilization and cost data from seven Supported Employment programs, in both rural and urban settings, which obtained fidelity scores in the range of 70-75 on the Supported Employment Fidelity Scale. Annual direct cost per client served ranged from $860 in New Hampshire to $6,793 in Indiana. State directors, as well as local mental health directors, may be interested in the benefits of a formal and comprehensive model that addresses both variability and uncertainty in the monitoring, planning, budgeting and quality improvement of EBPs within systems of care.

Statistical modeling, or simulation, is demonstrated as a method of planning and forecasting that allows system designers to be explicit with regard to the probabilities of given events and values used in making certain assumptions that can influence various outcomes. Using computer software, decision makers can enter any number of critical “input” variables, define one or more calculations as “output” variables, and write formulae that define the relationships among independent and dependent variables. The model provides an estimate of the probability that each outcome will occur. By changing the size or value of a variable, the decision-maker can simulate the probable effects of making specific “real-world” changes. This is more than “what-if” planning commonly done with most spreadsheet models. Stochastic simulation allows the planner to enter specific distributions of values and runs thousands of iterations to determine a distribution of possible outcomes. Therefore, the planner not only gets a calculated result, the planner can also determine the relative probability of a range of results (e.g. “I am 90% certain the annual cost per client served will range between $1,000 and $2,500, with a mean of $1,500.”)

Evidence Based Practices

In earlier stages of practice in the mental health field, the focus was on obtaining funding for services, followed by an era of focus on providing services in community settings rather than institutions. Although great strides were made, interventions alone have not changed outcomes for many persons with severe mental illness. The current era of mental health practice requires a focus on the content and quality of treatment offered. Quality and accountability involve adherence to EBP and fidelity to those specific program models that are shown to produce consistently cost effective results. Without model fidelity, an organization risks not achieving the positive outcomes demonstrated in the research. In fact, growing evidence finds that even some of the most popular and well-disseminated programs are not evidence-based and in fact can be counter productive (Goldman, et al., 2001).

EBPs utilize concepts that are empirically based. There must be valid and reliable research that has been replicated across diverse populations and settings with methodologies that are capable of being replicated in non-research settings. The strength of the scientific method is that it uses strictly defined, standardized procedures to determine how events are causally related. Using the scientific method also fosters recognition of the diversity of approaches involved in implementing programs and extracting data. Science based programs are theory driven, have program activities related to the theory and are reasonably well implemented and evaluated. From science based programs come those programs that are *promising*, and those that are demonstrated to be *effective. Effective* programs produce consistently positive outcomes and have been carefully implemented and evaluated. Generally this means that there have been at least three replicated studies that produce like outcomes. These *effective* programs are used to create *models* available for dissemination that generally incorporate technical assistance available from the developers.

<u>Evidence-Based Practices Drive Planning and Budgeting</u>
Many of the behavioral health services offered and paid for by State governments were historically based on services that grew out of de-institutionalization or were based on what were considered to be “essential” services relative to some consensus among federal and state mental health leaders. The job of planning and budgeting was often reliant on adding a component to an existing service or service array. State budgets seldom offered opportunities to develop totally new services that required a “zero-based” approach to planning and budgeting. With the emergence of EBPs backed by supporters and advocates, newer programs are more likely to be initiated and funded by reductions or total elimination of some of the historically funded but less effective services.

These Evidence-Based Programs (EBPs) offer new challenges because while they offer evidence for good outcomes and high levels of patient satisfaction, they may not always provide sufficiently detailed guidelines on operational expenses to allow straightforward planning and budgeting. In cases where EBPs are quite specific with regard to staff patterns and service intensity, such as Assertive Community Treatment Programs, the regional variability in the cost of most program items can present a significant challenge to States that wish to replicate these programs with some reasonable degree of fidelity (2, 10). Furthermore, due to the lack of certain credentialed professionals in some areas of the country, an EBP program may need to be modified with respect to its staffing pattern. Differences in the urban or rural placement of programs and the associated demographics, as well as associated differences in cost-of-living expenses can also call for a deviation from any empirically based cost standards.

Supported Employment

Employment provides a sense of purpose, a source of dignity and an opportunity to contribute to the community. Chronic unemployment leads to isolation, poverty, a diminished sense of self-esteem, perpetuation of homelessness and hinders recovery (6). Research has shown that the rate of unemployment for people with severe psychiatric disabilities is 85% or more (6, 8) and that 70% of adults with a serious mental illness desire work and that 50-60% of adults with a mental illness can be successful when working with supports. In fact 58% of those in supported employment obtained competitive employment compared to 21% in other programs (11).

Supported Employment was never intended to serve the typical vocational rehabilitation client. Supported Employment was created for those people with significant disabilities who could not obtain competitive employment through vocational rehabilitation services, who were not successful in competitive employment and who needed extended services and supports to achieve success. The goal was not to simply find jobs, but rather to find meaningful work. Supported Employment is defined as paid real and meaningful work in an integrated setting with ongoing support provided by an agency with expertise in finding employment for people with disabilities.

In the last decade, the concept of supported employment has become blurred by a plethora of models and approaches. It is the Individual Placement Model that has been most successful in enabling people to find meaningful work and sustain it within the local community. Supported Employment rests on the premise that real community jobs should be available to all people and that people should receive equal pay for equal work. Services must be designed to account for individual skills and preferences while accommodating the episodic nature of psychiatric illness. The supported employment program must offer support both off and on the job to be effective.

<u>Individual Placement Model For Competitive Employment</u>

People are not screened for work readiness in the Individual Placement Model. Services are available for anyone who says they want to work. Eligibility is based on consumer choice. Job search begins immediately after a consumer expresses interest in working. There is no pre-assessment or extensive training, immediate work experiences such as prevocational units or transitional employment, or sheltered workshops. Employment specialists use brief screening and self-report instruments for detection of substance abuse, domestic violence, housing concerns, health or learning disabilities that may require accommodation. Usually a case manager has this information and a review by the employment specialist facilitates the rapid development of supports. This means working with any individual whether or not they are in a stage of recovery that enables them to engage in full-time employment (6).

SE Model Fidelity

The Fidelity Scale lists the following items that must be accounted for in planning and budgeting:

- Caseload size: Employment specialists manage vocational caseloads of up to 25 clients.
- Vocational services staff: Employment specialists provide only vocational services.
- Vocational generalists: Each employment specialist carries out all phases of vocational service, including engagement, assessment, job placement, and follow along supports.
- Integration of rehabilitation with mental health treatment: Employment specialists are part of the mental health treatment teams with shared decision making. They attend regular treatment team meetings (not replaced by administrative meetings) and have frequent contact with treatment team members.
- Vocational unit: Employment specialists function as a unit rather than a group of practitioners. They have group supervision, share information, and help each other with cases.
- Zero exclusion criteria: No eligibility requirements such as job readiness, lack of substance abuse, no violent behavior, minimal intellectual functioning, and mild symptoms.
- Ongoing, work-based vocational assessment: Vocational assessment is an ongoing process based on work experiences in competitive jobs.
- Rapid search for competitive job: The search for competitive jobs occurs rapidly after program entry.
- Individualized job search: Employer contacts are based on clients' job preferences (relating to what they enjoy and their personal goals) and needs (including experience, ability, symptoms, and health, etc., and how they affect a good job and setting match) rather than the job market (i.e., what jobs are readily available).
- Diversity of jobs developed: Employment specialists provide job options that are in different settings.
- Permanence of jobs developed: Employment specialists provide competitive job options that have permanent status rather than temporary or time-limited status, e.g., Transitional Employment.
- Jobs as transitions: All jobs are viewed as positive experiences on the path of vocational growth and development. Employment specialists help clients end jobs when appropriate and then find new jobs.

- Follow-along supports: Individualized follow-along supports are provided to employer and client on a time-unlimited basis. Employer supports may include education and guidance. Client supports may include crisis intervention, job coaching, job counseling, job support groups, transportation, treatment or medication changes and network supports (friends/family).
- Community-based services: Vocational services such as engagement, job finding and follow-along supports are provided in natural community settings.
- Assertive engagement and outreach: assertive engagement and outreach (telephone, mail, and community visits) are conducted as needed (4).

Building Planning and Budgeting Models: An Example

Normally most planning and budgeting models are built using spreadsheet software, such as Excel. In designing and building a spreadsheet model the analyst usually makes a distinction between "input" variables and "output" variables. Input variables are the categories of data in the model that are mathematically inter-related through formulae in order to calculate the "output" variables of interest. The output variables depend on the purpose for which the model was built. The distinction between "input" and "output" variables is similar to the distinction made in research designs between "independent" and "dependent" variables. The former are presumed to be the operations or variables that can be controlled and modified by the researcher, while the dependent variables are the outcomes measured for each set of independent variables. For the State-level Director of Behavioral Health Services, the dependent variables or outputs are likely to be defined as:

- Total annual revenues and costs for the program,
- Cost per service unit, and
- Cost per service user.
- Functional client outcomes (e.g. number employed)

Within the spreadsheet, the planner/budget analyst is expected to enter the values of certain "input" variables, (e.g. number of clients/consumers expected to be served, their patterns of utilization intensity, the length of time clients will remain active in the program), or certain operational expenses and the costs of a specific staffing pattern necessary to maintain a capacity adequate to meet the estimated demand for services. Normally the planner/analyst will be expected to enter an <u>average</u> value when they believe the values could assume a range of possible values (e.g. the average length of time a client will remain active is 12 weeks.). These input values will be interrelated by formulae to calculate critical "output" variables (e.g. Total Cost, Cost per Client Served). Once the model is built, the planner can enter alternative values for the input variables to test their effects on the calculated output variables. However, if there are a large number of input variables, each with a number of possible values, it means there will be multiple combinations for the planner to test.

Consider a service system that has negotiated a risk-sharing contract with

Medicaid. The service delivery system will be pre-paid $30 per month for every Medicaid-eligible person in the community who may access these services, but the system is liable for all treatment expenses. The system needs a plan and a budget in order to provide some assurance that they will not lose money on this risk-sharing agreement. The system has access to historical information about the average values of key input variables for the Medicaid clients. The populated worksheet calculates an estimate of next year's total program costs as well as derivative indices like "average cost per eligible population member". Generally data is entered using last year's average values, usually with some up or down adjustments based on anticipated changes in how the program intends to operate, or perhaps based on new screening and authorization criteria to be used to accept patients, cost of living increases or other changes in the local environment that could influence program costs. A very simplified model of such calculations might look something like the spreadsheet illustration in the top half of Figure 1. The data in the lower half of Figure 1 illustrates what actually happened by the end of that year.

Figure 1: A Simple Model to Illustrate Basic Calculations

A. Initial Assumptions based on Historical Averages

Eligible Medicaid Population: 50,000

Program/Service	Average: Users per 1000 Eligibles	Average: Units per User	Average: Cost per Unit	Total Cost
Emergency Care	40	3	$ 250	$ 1,500,000
Inpatient Care	15	21	$ 600	$ 9,450,000
Partial Hospital	35	12	$ 125	$ 2,625,000
Outpatient Care	100	8	$ 85	$ 3,400,000
Total Clients/Consumers	9500		Total Costs	$ 16,975,000
Clients/Consumers per 1000 Eligible Members	190			

Annual Cost per Eligible Population Member	$ 339.50
Cost per Member per Month, (PMPM)	$ 28.29
PMPM Premium Payments from Risk-Sharing Contract with Medicaid	$ 30.00
Projected Total Revenue	$ 18,000,000
Projected Gross Savings (Loss)	$ 1,025,000

B. Actual Results in the Following Year

Eligible Medicaid Population: 40,000

Program/Service	Actual Rates of Utilzation and Unit Costs: Users per 1000 Eligibles	Units per User	Cost per Unit	Total Cost
Emergency Care	50	4	$ 250	$ 2,000,000
Inpatient Care	22	14	$ 600	$ 7,392,000
Partial Hospital	45	14	$ 150	$ 3,780,000
Outpatient Care	90	8	$ 95	$ 2,736,000
Total Clients/Consumers	10350		Total Costs	$ 15,908,000
Clients/Consumers per 1000 Eligible Members	259			

Annual Cost per Eligible Population Member	$ 397.70
Cost per Member per Month, (PMPM)	$ 33.14
PMPM Premium Payments from Risk-Sharing Contract with Medicaid	$ 30.00
Actual Total Revenue	$ 14,400,000
Actual Gross Savings (Loss)	$ (1,508,000)

Clearly the actual outcome was much different than forecasted. What could have gone wrong? Why were the historically derived average values unreliable as forecasts for the future?

- Welfare reform reduced the number of AFDC-eligible persons but not the number of SSI-eligible persons, so the actual number of SSI as a percentage of total eligible persons goes from 20 percent to 25 percent and the total eligible population was changed. Furthermore the total size of the population, on which revenues were based, dropped to 40,000 and the prepaid revenues at the rate of $30 PMPM, dropped from a planned $18,000,000 to only $14,400,000.

- Although the total population of eligible members dropped, the relative increase in the mix of SSI-eligible persons led to a client case-mix of relatively higher severity. While the projected number of clients per 1000 eligible members was only 190, the actual rate of clients per 1000 eligible members was 259.

- The relatively greater number of SSI-Eligible members led to increased use (more users) of emergency care coupled with more frequent emergency visits within the year (4 times instead of 3).

- There were more readmissions to the hospital so the effective admit rate is 22/1000 lives and not 15/1000 lives.

- The average length of stay (LOS) for inpatient was reduced from the planned 21 days to 14 days but the SSI-Eligible clients used more partial hospital care and had more sessions.

- Due to the population's greater difficulty in maintaining levels of compliance with scheduled appointments, your outpatient staff's productivity level was reduced and your effective cost per unit of counseling went up to $95 per hour.

Since these actual events were realized, the actual Total Cost was $15,908,000, or $33.14 PMPM, compared with the anticipated Total Cost of $16,975,000, or $28.29 on a PMPM basis. Although the Total Costs were lower, Total Revenues were also reduced because the population became smaller, and the utilization rates per 1000 Eligible Members increased, with the greater increases in the most expensive programs. The total number of clients across all programs increased from the projected 9,500 to 10,350.

In summary, this difference in the outputs or outcomes occurred because the input variables did, in fact, vary from the expected averages. Using averages in standard spreadsheets, even with estimated adjustments, means that

- Rare events with highly costly outcomes are more likely to be over-estimated in their probability.
- High frequency events with low to moderate costs are generally under-estimated.
- A large number of moderately costly events (i.e. cost per event) will, in the aggregate, cost more than the aggregate cost represented by the sum of a very small number of high-cost events.

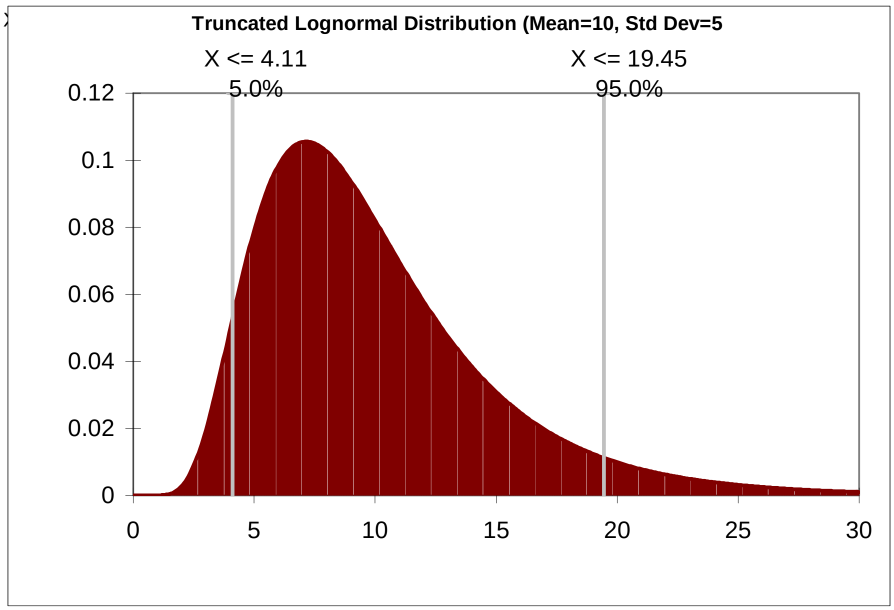
Furthermore, one cannot accommodate or anticipate all of the potential values and variable combinations such as increases or decreases in access rates, service units, and unit costs across all the programs in the system. In the above example, there were three input variables for each of four programs, or 12 total input variables, and the size of the eligible population making a total of 13. If each variable's value is independent of every other variable's value (which in this case is not a totally correct assumption (e.g. reductions in hospital length of stay could increase number of emergency program clients as well as the average number of emergency visits), and each input variable could take on as few as only five possible values (a very conservative assumption), there are 5 to the 13th power number of combinations and values for the output variables that are possible or 1,220,703,125 possible combinations of input values, producing an equal number of possible output values, like Total Cost and PMPM cost calculations. While some of these possible combinations will be identical because increases in one variable can be offset by decreases in another, one is still left with testing a huge number of possible combinations. And after testing even as few as 1,000,000 combinations, can we be assured that the analyst selected the most likely values from among the five possibilities for each of the 13 input variables? Worst yet, which one of these 1,000,000 outcomes is the most likely to occur?

The "Flaw of Averages"

Whenever most people are told of an average value, (e.g. average number of outpatient visits per closed outpatient episode) they generally do NOT consider the impact of variation around that average value. If they do consider possible variation, they generally assume a symmetrical distribution of variation around the average value. This type of distribution is called a "normal" distribution, and it is the distribution that most of us implicitly assume when we think about variability. Often the claim is that "variations to the high side will be off-set by variations to the low side". However, most distributions of services and costs in the health and human services industry are NOT normally distributed. Rather, they are "skewed" distributions, wherein the average value is greater than the "modal" (i.e. most likely value) because there are a small number of very high values that "drag up" the average. In other words, there are small probabilities of very large values and very large probabilities of moderate or small values. In some cases, as when plotting the distribution of "total cost per client per year", it is common to observe a "Pareto", or "80-20" distribution, where only 20% of the clients/consumers will absorb as much as 80% of all the resources.

See Figure 2 below for an illustration of a lognormal distribution, commonly found in length of stay (LOS) data. This distribution reflects data on the LOS in an inpatient program from only discharged cases, whose admissions and discharges were within the fiscal year. This distribution has a mean value of 10 and a standard deviation of only 5. It is further truncated so the minimum value can only be 1 (i.e. no zero length of stays) and the maximum value is 365 (the maximum LOS in any given fiscal year).

Figure 2: A Lognormal Distribution with Mean=10 and a Standard Deviation=5.



Statistical Parameter	Value
Minimum	1.00
Maximum	364.45
Mean	10.00
Mode	7.12
Median	8.94
Std. Dev	5.00
Left X	4.11
Left P	0.05
Right X	19.45
Right P	0.95
Diff. X	15.34
Diff. P	0.90

To keep the graphic presentation within the limits of the page width, the distribution appears to end at 30 days, but in reality the “X” axis could be extended out to 365 days, showing the probability of a very small number of cases with that LOS. While the mean LOS of this distribution is 10, the modal length of stay was only 7.12. The LOS that was most common among all discharges was 3 days less than the “average” LOS. And the median LOS was almost 9 days, again less than the average LOS. Only 5% of all discharges had a LOS equal to or greater than 19.45 days and 5% had a LOS of 4.11 or less. Ninety percent of all cases had a LOS between 4.11 days and 19.45 days. This distribution provides more information than the average value would provide.

Table 1 lists common types of distributions characteristic of input variables used in common budgeting attempts. Most of these distributions have a positive skew as seen in Figure 2 above. These high values will “drag up” the average so that the average value will be higher than either the modal value (the value most cases in the distribution have) or the median value, (the value that separates the lowest 50% of all values from the upper 50% of all values). More severe distortions are introduced when average values for input distributions have nonsymmetrical, right-skewed distributions.

Table 1: Common Types of Distributions Characterizing Cost and Utilization Data in Health and Human Services Data Sets

Normal:	Distribution of characteristics of a population (height, weight); size of quantities that are the sum of other quantities. Normal distributions are symmetrical; median = mode = mean of all possible (n) values Variance = sum of the (deviations from the mean) squared divided by (n-1)
Gamma: **See Pert also**	Time to complete some task, such as building a facility, servicing a request, completing a treatment episode Mean = (Alpha x Beta) Beta = Mean - Mode Mode = Beta (alpha-1) Alpha = Mean/Beta
Pareto	A highly right skewed distribution of entities based on their value on some variable elative to other entities' values. Typically, a very small number of the entities account for the dominant share of all the costs or resources (e.g. the 80-20 rule). This distribution describing such outputs as the annual cost per case served.
Poisson:	The rate of rare events that occur in a given unit of time. (Number of annual admissions into a program; number of accidents in an intersection, incidence rate of many diseases) mean = lambda positive integers only
Lognormal:	Represents quantities that are the product of a large number of other quantities. (e.g. cost per inpatient episode (LOS X Cost per Day), distribution of physical quantities in nature (e.g.size of oil fields)
Triangular:	Distributionfor estimates when actual data is missing. Expected, least likely, and most likely, possible values
Binomial:	Distribution of outcomes when there are only two, mutually exclusive possibilities n = number of events p = positive outcomes q= (1-p) negative outcomes mean = np Std Dev = square root of (npq)
Beta-Subjective	Used for general distributions and modeling in the absence of well-fitted empirical distributions.

This article does not allow for a full discussion of the various types of frequency distributions that can represent the variation around the mean value of common input variables, such as Clients/1000, Units per Client, Cost per Episode, Annual Cost per Client per year, Time in Program, etc. Mean distributions can be extremely skewed, (i.e. Pareto distributions) and others less skewed or nearly normal (e.g. Poisson distributions). Furthermore, the reliability of the average, (i.e. the extent to which the average can truly represent the total distribution of values) of some of these distributions is a function of the total sample size, the number of clients or episodes, etc. reflected in each distribution. Given very large samples of data, with a very large number of episodes or service units, the distribution of the average values from skewed distributions will begin to approach "normality" in their shape.

Because variation is not likely to be symmetrical, one should not look only at average values to best characterize the "central tendency" of input variables. It is critical to consider the shapes of the distribution in your historical data sets for fully understanding and measuring the possible variation above or below the average that are used as input variables for estimating utilization, costs, or revenues.

Stochastic (Probabilistic) Models for Planning and Budgeting

To assist States in planning and budgeting for EBPs, formal mathematical models should be utilized that allow the planners to enter distributions, rather than averages, as input variables to arrive at estimates of utilization or costs. When the values of input variables are seen as probability distributions, rather than some fixed number based on an "average" or some other estimation method, the model is said to be a stochastic (probabilistic) model. When historical data is missing, or incomplete, or no longer valid due to many changes from that point of history to the current period, stochastic models allow the analysts to reflect their real uncertainty about what values to enter by allowing them to enter distributions to use as estimates for important input values, such as utilization intensity patterns as well as certain types of expenses using a distribution rather than entering a single value.

Sampling Input Values and Producing a Distribution of Possible Output Values

For purposes of planning future EBPs and their potential costs and revenues it is important to understand the inter-relationships among important input and output variables. The term "variable" means that factors can indeed differ across some range of possible values. Furthermore, as seen above, "average" values may be misleading if calculated on the basis of a skewed distribution of possibilities. An appropriate planning and budgeting model will incorporate and mathematically inter-relate these critical input variables (i.e. distributions) and allow the user to run a simulation (e.g. have the spreadsheet re-calculate itself many times - 1,000 or more), each time sampling possible values of the input variables and producing 1,000 or more potential values of the "output" variables. In this manner a frequency distribution of possible output values, (e.g. the upper and lower limits of costs and revenues, the cost value that is most likely, based on the relative probability of the possible values that the input variables could assume.) is created.

The values chosen for each input distribution on each iteration are not randomly chosen but sampled from the distribution based on the relative likelihood of each value in that distribution. Values in and around the modal value will be chosen (sampled) relatively more frequently than extremely high but unlikely values in the tails of the distribution. Unlike the human mind, that would select values based on subjective biases, (e.g. over-sample unlikely high values, and under-sample more likely moderate or low values); the software used in the model will choose values based on their relative likelihood (probability) of occurring. The model does not over-sample high values or under-sample low values. Given that there are 1,000 iterations of possible input values, the model will have calculated and saved 1,000 “answers” or possible values for the critical output variables. These 1,000 values of any given output variable can be represented as a distribution, providing the analyst with a graphic representation of the likelihood of any given output value, as well as the range of values that are most likely to occur.

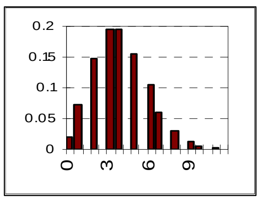
For example, imagine a spreadsheet model used to estimate the cost of inpatient episodes being recalculated 1,000 times. Each time an iteration occurs; a different value for each input variable is sampled from the distribution and used to calculate a new value for the important output variables. This simple model might be used by a managed care organization (MCO) that purchases inpatient services from a variety of hospitals throughout a large geographical area where the managed care membership lives. Each hospital has negotiated a different per-diem rate. The MCO needs to know if its inpatient cost PMPM is likely to exceed the PMPM payments they are collecting to manage inpatient psychiatric hospital admissions.

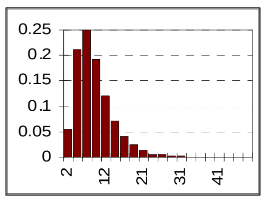
The first input variable is the admissions rate, expressed as a rate per 1000 members. The size of the “membership” is 100,000 or 100 cohorts of 1,000. A poisson distribution is generally used to represent admission rates. This particular poisson distribution has a mean and variance equal to 4, or 4 admissions per year per each cohort of 1,000 members. A rate of 4 per I,000 would equate to 400 annual admissions, since there are 100 cohorts of 1,000. The second input variable is the LOS for each episode of care. This particular lognormal distribution has a mean of 10 and a standard deviation of 5. The third input variable is the cost per day. Based on the contracts it has signed, the MCO could readily determine that this distribution of per-diem rates was nearly symmetrical with a mean of $400 and a standard deviation of $50. Therefore a normal distribution was used to represent the variability of this third input variable.

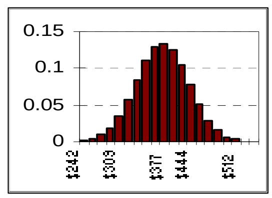
The formula that inter-relates these three inputs to the output variable of interest; cost expressed as a PMPM cost is: [(rate of admissions times the membership size divided by 1,000) times LOS times cost per day] = Total Cost / divided by the (membership size and divided again by 12) to get a monthly cost per member. Figure 3 illustrates how multiply sampled input values would create multiple values of the desired output: Cost for Inpatient Services on a PMPM basis.

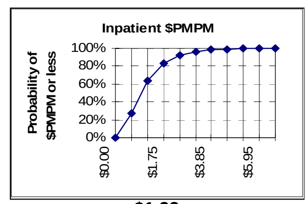
Figure 3: Illustration of How Samples of Inputs Create Multiple "Outputs" or Outcomes

Example Only: Inpatient Cost | **Number of Lives = 100,000**









Iteration	Number of Users per 1000 Poisson (4)	Average Units per User Lognormal (10,5)	Average Cost per Unit Normal (400,50)	Estimated Inpatient Cost ($PMPM) [(F4/1000)*B10*D10*F10]/F4/12
1	4	10	$400	$1.33
2	2	12.9	$410	$0.88
3	6	9.3	$429	$1.99
4	1	5.9	$326	$0.16
5	2	10.9	$407	$0.74
6	5	7.7	$442	$1.42
7	3	14.1	$396	$1.40
8	1	12.4	$365	$0.38
9	3	9.8	$429	$1.05
10	0	11.8	$399	$0.00
11	2	7.5	$368	$0.46
12	2	5.5	$435	$0.40
13	4	6.7	$409	$0.92
14	8	9.9	$330	$2.17
15	2	5.1	$418	$0.36
16	4	9.5	$463	$1.47
	:	:	:	:
	:	:	:	:
	:	:	:	:
1000	3	8.7	$425	$0.92

Each row represents a particular iteration of the model, producing a unique combination of three input values that were sampled. To conserve space, only the first 16 iterations are shown along with the last one of 1,000 iterations. Notice in the first column, that the values under the Poisson distribution are always integers and tend to cluster around the distribution mean rate of 4 per 100 per year. However, on the 14th iteration the sample included a less likely but high value of 8 admissions per 1000 members per year. This particular iteration also produced the highest PMPM cost of $2.17, suggesting the variability in this input is highly correlated with the variability in the output. In the second column, you will notice that sampled values were the most likely to cluster around the modal value of 7 and the most extreme value sampled for LOS was on iteration 7 with a value of 14.1. The per-diem rates that were sampled in column three were clustered around the mean of $400 dollars. Because rate is a symmetrical value, values are sampled below the mean as often as values above the mean.

The crucial output variable in column four, cost per member per month, had 1,000 possible values which were used to form a distribution and that distribution was then converted to a cumulative ascending distribution so the level of probability for any PMPM cost could be easily read from the chart. For example, it is 60% likely that the PMPM value will be $1.75 or less. Conversely, one could say there is a 40% chance it could be $1.75 or more. In either case the MCO could quickly assess if their cost is likely to be more or less than the revenues they are being paid to manage inpatient psychiatric care.

This model could also be run in a manner that would identify which of the three input variables is the most important to manage and control in order to reduce the uncertainty or variability of the output variable. Uncertainty or high variability is not good for an MCO’s ongoing success, nor helpful in the implementation of EBPs. Manufacturing companies for example, equate low variability with high quality because when variability is low each product is being produced within very narrow standards. In addition, the model can be run to identify the minimum values that each of the three input variables must have in order to achieve a given level of certainty that the output variable will be equal or less than a specified output value. For example you could run the model to simulate what the three input variables must be in order for to be 50% or more certain that the PMPM cost will not exceed $1.50. The three input variables are connected by a single formula that is used to calculate the output variable’s value. Most real models are much more complicated and may have several hundred input variables that are inter-related through hundreds of different formulae. An illustration of such stochastic modeling for an EBP Supported Employment Program is presented later in this document.

<u>Important Input Variables For EBPs</u>
The three most important input variables that will drive the cost of any intervention or EBP service within a defined eligible population are:

1. The **rate of entry** by eligible persons into the program (Participants, Consumers, Client or **“Users”**);
2. The **Intensity** of services, defined by the **Units of Service** within a **Period of Time**, provided to each client/consumer;
3. The cost of these units of service (**Cost per Unit of Service**).

The values of these critical factors are being driven by other variables such as case-mix severity, the number of stratified cohorts being separately tracked, staffing pattern (type of staff by credentials and salary plus fringe benefit costs), staff productive levels, the days of the year the program is operational, etc. The cost of a unit of service delivered by a program will depend on variables such as the following:

- The employee's salary and fringe-benefit costs
- The average hours in the year that the employee spends in "productive" service (i.e., the cost of a productive hour of service)
- The cost of direct expenses necessary to establish that particular service (e.g., equipment, supplies)
- All variable costs, (supplies, gasoline for travel, consumer meals), expenses that will depend on service volume and will further depend on the kind of service being planned (e.g. mobile staff travel to the clients, or clients come to a facility, like a residential program that operates 365 days per year on a 24hours/7 days per week)
- The cost of General and Administrative expenses necessary to support the general organization and that particular service (e.g., insurance, utilities, administrators' salaries)

These variables may be further broken down and related to one another through a formula such as the following:

Cost of a Staff Person's Hour of Service = [(Salary) + (Fringe Benefits) + (Share of Direct Expenses) + (Share of G&A Expenses) / Total Productive Hours]

Therefore, because the cost of these primary variables are driven by the values of even more nested secondary variables, the distribution of values for each of these three primary input variables is usually distributed across a wide range of possible values.

The Importance of Program Capacity

Because Users and Units per User are important variables and both of these variables are generally influenced by any limits on the number of clients to be served, a fourth critical variable becomes program Capacity. It is important to consider whether capacity is fixed, or potentially variable. A good example of a fixed capacity program is an inpatient hospital unit or any other type of residential program. There is an upper limit on the number of beds so there is a fixed capacity. However, the number of unique persons to be served in such a fixed capacity program can vary, depending upon the LOS or "turnover rate". If capacity is defined as the number of clients served per year, then a residential program could be said to have a variable capacity.

A variable capacity program is one where the supply of providers is easily expanded or diminished to match the growing or shrinking demand by consumers of the service. For example, if all units are reimbursed by other parties (another level of government or the consumers themselves), and there is a large number of providers who could be hired or paid as contracted providers, a program's capacity could grow or shrink as demand dictates. Historically, most behavioral services paid for by State governments were designed as fixed capacity programs. In State reform efforts for managing community care and under managed care waivers, services are reimbursed on a variable basis, paying only when services are duly authorized and used by the consumer. The important questions for State government become:

- How much capacity is necessary to meet appropriate demand?
- How much capacity can I afford to pay for, relative to the capacity necessary to meet appropriate demand?"
- Is there any way to design the program in order to make the capacity somewhat variable, with lower fixed costs?"
- What is the effect of client choice of provider, and what are the necessary numbers of cases each provider must serve to be viable?

The issue of capacity is further complicated by how it is defined. Some may think of it as the number of persons who can be actively served on any given day, while others may define it as the number of unique individuals that can be served in a given year. While we prefer the latter definition, it also begs the question: how do you define a person as "served"? Does one person using one session in a program count as equivalent to a person who completes 12 sessions? Therefore, to measure the number of persons served in a year, one must have a definition of what minimum level of service consumption defines a person as being served. One can also report on such measures as total client-days of service.

Capacity is also determined by the number of days in the year that the program is operational. An after-school program that can handle 50 children for up to 4 hours on any given day, but is operational for only 5 days per week for 36 weeks (i.e. the school year), has less capacity than a day treatment program that is operational 8 hours per day, 5 days a week for 52 weeks.

Staff productivity can influence program capacity. A program where each clinician has 75% of their payroll hours used for direct services, has more capacity that one where clinicians have only 50% of their payroll hours spent doing direct services.

Talking about capacity is almost equivalent to talking about cost. The degree to which capacity is fixed, variable, or blended is dependent upon the overall program design. For example, a case management service may provide an individual office and computer for each case manager, so fixed costs are a function of the number of case managers, but costs are variable to the extent that their staffing and client numbers may grow or shrink. On the other hand, costs may be lower, but more fixed, in a case management program where a team of case managers share a common caseload and are expected to work as a team, where all team members back up other team members sharing a common caseload. The fixed capacity costs are then likely to be based on case managers working in a large common space, sharing certain common equipment. The large space and shared equipment is a fixed cost as long as the number of case managers never exceeds the capacity of the common space or predefined “team”. This pattern is, in fact, the case for the EBP of Supported Employment.

Capacity planning can be troublesome. Programs with too much capacity run the risk of being labeled as under-utilized and inefficient, with high costs per person served. If too little capacity is planned and demand is high the State may be blamed for not providing a level of service adequate to meet the clients’ needs and expectations. When clients are allowed a choice of provider, careful planning for the number of providers necessary to meet choice requirements is critical. Certain programs cannot operate without a sufficiently high client base. In short, capacity planning can be a lose-lose type of activity. Generally speaking, designing programs to have variable capacity, and variable cost, are more likely to be efficient and satisfactory to consumers, than are programs with fixed capacity that must limit access when demands exceed capacity and generally do not offer a choice of provider. Another phenomenon seen in fixed capacity programs is that staff members are more likely to maintain caseloads to avoid client turnover and additional work, thus creating longer lengths of stay per case and fewer persons served per year than would occur in variable capacity programs.

Other Variables Affecting EBPs

Especially when considering EBPs, other variables need to be considered such as **start-up costs** that are likely to drive up unit costs due to training, the cost of maintaining required staff to client ratios, as well as the on-going cost of monitoring and maintaining **compliance to fidelity standards**.

Because other program fixed and variable costs are often budgeted by **allocation** per Full Time Equivalent staff (FTEs) other factors are nested in the capacity variable. Costs can be classified as “ongoing” or “one-time”. General accounting standards usually convert most one-time or start-up costs into ongoing annual costs by amortorizing one-time costs over variable time periods. The cost of a building or buying a facility might be depreciated over 20 years, so 5% of this initial cost is considered as an expense for a one-year fiscal period, for each of the next 20 fiscal years. Some equipment may be depreciated over shorter time periods, while supplies and other equipment may be considered as expenses for the year in which they were purchased. In most service programs the primary ongoing costs are based on staff salaries and benefits, so the capacity questions often boil down to staffing pattern questions. As most administrators know, staffing costs can generally represent 60% to 90% of the total program budget. Furthermore, most staff costs are considered fixed in the short run of one-year

Many **characteristics of a population** of potential service users may affect their service use patterns. Examples include: the ratio of males to females; the number of each who fall into various age groupings; and the number who live in single-parent homes, have little or no education or skills training, or lack transportation or other social supports, all of which are critical in the Supported Employment world. However, when these characteristics are known, reimbursement systems can be tailored to ensure that reimbursement levels are adequately addressed.

A **pattern of use** among members in the population creates a combination of factors that directly affect financial risks. In the Supported Employment example, how many persons will seek employment services? What kind of job will they seek? Will the placement be effective or will they have to find another? If given a choice, will they go to high-cost or lower-cost providers? Historical utilization data or data from secondary sources may or may not answer these questions.

When members have **choice** among competing provider systems, they will be influenced by factors such as ease of access, the provider's reputation, perceptions of quality, or past experience with the particular system. Those who choose a particular provider system may be healthier or sicker than the average Member within the defined population. The direction of this bias toward adverse or favorable selection cannot be specified in advance for all conditions. Being selected may be a good or a bad outcome for a provider, depending on the nature of the risk or payment arrangement.

Table 2 provides a summary of how capacity (number of persons served in a year) and cost are a function of the type and intensity of service being offered.

Table 2: Factors Affecting Program Capacity and Costs

One-on-One Programs (e.g. Outpatient)
- Number of Direct Service Staff and Their Time Available for Direct Service
- Point-in-Time Active Caseload per staff
- Rate of "No-Shows"
- Average Visits per Client

Group Programs
- Number of "slots" that are available each day
- Days of the Year the Program is Operational
- Turnover Rate Among Active Clients

Emergency Programs
- Direct Service Staff and Their Time Available for Service
- Time per Intervention
- Days and Hours per Day Available for Service (# of Shifts)
- Days of the Year the Program is Operational

Case Management
- Direct Service Staff and Their Time Available for Service
- Point-in-Time Caseload per Staff
- Turnover rate Among Persons Active in the Caseload

Residential
- Number of beds available per day
- Number of days per year in operation
- Turnover Rate among Bed Occupants (i.e. Average Length of Stay)

Designing and Building Stochastic Models

Because of the different variables that drive capacity, a stochastic model for planning and budgeting is designed differently depending upon the type of service being considered. The differences in design, and how the input variables are mathematically connected by formulae, are driven by how program capacity can create upper limits on the number of persons to be served in a given year. The model's design will also depend upon whether it is to be used to assess risk, or simply cost, and whether it is a model of a single program or a set of programs spanning a large State system or network of contracted providers and whether the organization using the model owns and operates the services directly, as in some States, or whether they purchase these services for citizens through contracts or grants provided to community-based providers.

Understanding that capacity is an important yet difficult variable to define and measure, we can now summarize in Table 3 the critical formulae necessary for a stochastic model to be used for planning and budgeting. The following formulae assume that the model is used for a fixed time period, such as a fiscal or calendar year.

Table 3: Critical Formulae for Understanding Program Costs

1. Total Cost = (Users) X (Units/User) X (Cost/Unit)

2. Total Costs are a Combination of Fixed and Variable Costs

Using algebra

2. Cost/Unit = (Fixed + Variable Costs) / [(Users) X (Units/User)]

3. Users are a function of Capacity

4. Units per User (i.e. Service Intensity) is a Function of Planned Service Pattern

5. Variable costs are a function of the Number of Users

6. Fixed costs do not vary (during the year) as a function of Users

The processes for designing and building stochastic models are very similar to the process for building a house. Some models are highly customized to suit the specific needs of the person who will use it. The model, like the building, will have a certain size or scope and a certain number of “sections” or “modules”, just as a building may have a certain size and a certain room layout. Some persons want a model to represent a single program while others may want a model that represents an entire system of care, with many different programs, being offered by different providers within a contracted network. Some models are designed to estimate financial risk under various risk-sharing arrangements (payment per member per month - PMPM - or case rates) while other models may be built to get a measure of cost per service unit. Broskowski and Smith (1) published the results of models designed to determine the additional PMPM cost to HMOs if they were to include in their scope of benefits one or more of six behavioral healthcare interventions shown to have empirically-based evidence of good health outcomes. Once designed, it is up to the analysts to “populate” the model with distributions and values that are representative of their system of care just as a blueprint may reflect the design of a house, but leave it up to the owner to populate the various rooms with their own specific furniture and accessories.

Some models are not customized but designed as “generic” models that are flexible enough to be used by a range of different users. Generic models are like model homes. There may be only two or three alternative designs that the builder has found will meet the various and variable needs of a wide variety of people to whom he is marketing. For example, the generic model might have a separate tab sheet for each of 20 different types of programs. One user may use all 20 templates to estimate the cost for a large system of care, while another may only use 5 of the tabsheets because that system of care is much smaller.

Generic models can also be designed with a variety of “switches” that allow flexible alternatives for different users. For example, a generic model with 20 programs may have a switch for each program, to check-off whether the program is owned and operated by the payer, or purchased from a contracted vendor. If switched to “own and operate”, the model asks the analyst to enter budgeted expenses for staff and other expense items. When switched to “purchased”, the model asks for the prices per unit of service contracted, or the total aggregate cost of the performance contract if it is not being structured as a fee-for-service (FFS) payment mechanism.

This model is built to illustrate the EBP called Support Employment and is not necessarily the same type of model one would build to estimate cost and utilization for certain other programs in a system of care.

Controllable and Uncontrollable Sources of Variation

In designing the model it is important to select the right distribution to use for each variable, and to understand the underlying factors that are “driving” the variation. Many factors may be managed or controlled in order to reduce the variation. For example, formal algorithms for triaging clients to service programs, and formalized treatment protocols can reduce variation due to independent decisions made by clinicians. In fact, EBPs are a key to reducing variability in services and outcomes. There are some factors that cannot be easily managed and changed in the short run to reduce variation. For example, factors related to characteristics associated with a particular population of eligible users may not be easily changed (poverty, underlying prevalence of disease, culturally learned values about treatment use, etc.).

Another important source of utilization and cost variation can be traced to the payer’s requirements for unique or specialized services for individual users. The more customized features that are required and provided, the higher the cost of the overall system. This risk has been addressed in general health care systems by trying to reduce treatment variations that are based on providers’ or patients’ individual preferences, rather than their demonstrated effects on treatment outcomes. But increased sensitivity to individual or subpopulation characteristics also means increased variation and its consequence; increased service costs. For example, if direct service providers must be multi-lingual, or if they must support multiple access points in a neighborhood, then costs will be greater. In considering costs, providers should be alert to program requirements of the payer that can introduce increased variation in the estimated utilization or unit cost factors.

There are a large number of program design issues that will affect cost in EBPs (UM requirements, grievance and appeal systems, the extent to which Person Centered Planning is implemented and used, consumer choice of provider and any additional staff required to monitor compliance with fidelity standards.

A proper model must take into account that critical input variables can vary throughout the year by taking on a different value in any given month or a different value for each type of client/consumer served. Some variation may be due to the discretionary behavior of managers and staff and lack of knowledge or failure to follow a well-document treatment guideline or protocol, while other variation cannot be controlled because it exists in the environment and affects the program in uncontrollable ways (e.g. client/consumer variability in severity, weather patterns affecting heating/cooling/plowing costs, price of fuel affecting transportation costs for mobile service units, etc.) Modeling tries to distinguish these two types of variables by using separate terms: variability and uncertainty. A famous statistician, Sir David Cox, addressed the distinction when he said: *"**Variability** is a phenomenon in the physical world to be measured, analyzed and where appropriate, explained. By contrast, **uncertainty** is an aspect of knowledge"* (or lack thereof).

These two types of variability, due to uncontrollable but perhaps predictable chance, and controllable uncertainty, based on managerial and staff knowledge and personal, discretionary performance have to be considered and addressed in planning and budgeting models. To account for both types of variation, computer simulation is used to account for all possible values, and a sufficient number of iterations of the model are run to allow for the possibility of most input values and the associated values of key output variables.

Variation Due to Chance ("Probabilistic Variability")

The first type of variability, called "stochastic" or "probabilistic" variability, is the effect of chance and uncontrollable variation. Examples of stochastic variability include the percentage of eligible individuals who will present themselves for treatment or the severity of their condition. Some programs may try to control or limit such variability by having admission standards that reduce severity variation or admission rules that limit the number of persons who are given access to the program.

Often organizations can use historical data that provide a reliable average upon which they can estimate future use demands. However, if the program is totally new and historical data is lacking, as in the case of many EBPs, the model must depend on reasonable estimates, and offer the user the opportunity to express his/her degree of uncertainty by entering the parameters of a distribution that describes all possible values.

For example, in estimating the variable called "length of stay" the model may call for the entry of the following parameters that can be used to describe a distribution:

- Minimum possible value, (e.g. 1 day)
- Most likely value, or the mode of the distribution of all possible values, (e.g. the value that most clients will have is 4 days)
- Maximum, or greatest possible value (e.g. 90 days)
- Average value – the sum of all actual LOS values divided by the number of persons admitted (e.g. 27 days)
- Median value - that value which splits the distribution of all client values in half, with 50% being below the median and 50% being above the median, (e.g. half the clients stayed 10 days or less, and half stayed 11 days or more)

In the absence of historical data, the person using the model is forced to use estimates. These estimates may be based on program standards, past experience in similar programs, or goals set by the program's leadership. When using EBPs the model is more easily based on data from replicated research studies. Still, variability is created because of knowable variation but variation that is not easy to control. For example, in an urban area of the State, nurses' salaries are known but they are higher than salaries in rural areas, which are also known. What salary level should the model assume? Fortunately such knowable but uncontrollable variation can be handled through simulation tables. For example, on the first set of 1,000 iterations, the urban value will be used, but the same constant value will be used on each of the 1,000 iterations. On the next set of 1,000 iterations the single rural value is used for all iterations. The model will thereby produce two sets of results, one for urban and one for rural communities.

Variation Due to "Unknowns"

The second type of variability is due to the uncertain knowledge of the modeler. Such uncertainty represents the model builder's *lack of knowledge* about the actual possible values of some variables. Although the model builder may lack knowledge about the proper values of certain variables, their values are readily known or controlled by the system operators. For example, in a generic model for all EBPs the salary level and fringe-benefit costs of staff in different types of EBPs may be considered "unknown" variables in the model, not because these values fluctuate or vary within any given EBP, but because the model builder is not certain what values would fairly represent the universe of all EBPs.

Practical Design Issues

As noted earlier, both customized and generic stochastic models of behavioral health programs and systems of programs are designed in a modular fashion, so variables of a similar domain are organized on a single tabsheet in a muti-tabsheet workbook. The tabsheets are organized going from a left-to-right sequence, prompting the user to enter first variables first. The final tabsheets are places where key performance indicators and cost summaries, the most important outcomes of interest to the model’s users, are located.

The design of the mathematical model is similar in structure, the number of different service programs in the single organization or within a larger network of providers, to the real world system it is designed to simulate. The model can also be designed to reflect the differential results on key indices for different cohorts or sub-categories of clients within that same organization or network, such as adults versus children, or mental health users versus substance abuse users.

The model begins with a “**Population**” tabsheet where the user is asked to enter population and sub-population estimates. Population in the model refers to all possible persons in the environment that could have access to the program if they so choose. Therefore, the model may be limited to just adults with Medicaid eligibility, or children within the child-welfare system. In fact, most models are designed to have more than one population subcategory, making it easier to use epidemiological estimates of the “prevalence” of that specific range of conditions the service program or organization or network is designed to serve. For each population subcategory the model allows the user to enter separate values, from which it is then possible to enter estimates of the total number of possible users that might access the program, organization or network.

The next commonly used tab sheet is labeled **Clients**. Here the user enters the percentages of the population subcategories that are expected to become active clients within the program or system of care. One may also enter separate estimations for the percentage of total clients that will fall into alternative cohorts, based on severity grouping, or gender and age combinations.

If the model covers more than one program, as is commonly the case, the next tabsheet is labeled something like **“Triage” or “Assignment**”. On this tabsheet the user is asked to allocate the percentage of all clients that will be entering each of the multiple programs (provided you are using the model for more than one program). In this case, the sum of the percentages is greater than 100% because clients usually enter more than one program in a given year.

Each model has a **“Capacity”** tabsheet that can calculate the number of clients that could be served by a given program in a given year based on the capacity and average use of that program on any given day. The **“Triage”** tabsheet may be used to estimate what percentage of each client cohort will be served in each service program, so that clients can be placed, if necessary, in more than one program. The numbers of expected users of each program can be compared against the **“Capacity”** of each program and can flag those program assignments that go over a given program's allowable capacity.

To accommodate systems of care that will pay for services on a fee-for-service basis, there is a separate tabsheet called **“Prices”** This tab sheet incorporates a table that provides the unit price to be paid for each type of unit within each program, (some programs may provide more that one type of service unit).

On another tabsheet called **“Staff”** the user can specify the staffing pattern called for in each program in the model, the staff’s productivity levels, and use lookup tables to assign salary and fringe benefits for each category of professional (i.e. psychiatrist, social worker, psychologist, etc.) or support (i.e. van driver, maintenance worker, clerical support, etc.) personnel.

Administration is handled as a separate type of **“program”** with its own tabsheet for staffing and budget expenses. The staffing pattern might include a wide array of various administrative and support positions that are not specific to any single program. Managers and other administrative personnel who are specific to a given program will be placed in that program’s staffing sheet. Capital expenses and general and administrative start-up as well as on-going expenses that will serve all programs in the system (e.g. phone system, information system can be allocated based on amortization over a defined time period, or the model can offer the user an opportunity to choose alternatives for outright purchase, lease, or financing of these capital items.

In order to determine a comprehensive cost for each program, one that includes direct and indirect expenses, general and administrative (G&A) expenses are then “allocated” to separate direct service programs based on one or more rules defined by the model’s users, (e.g. allocate G&A expenses in proportion to the total direct and variable expenses of each program, allocate by service or client volume, etc.). It is also possible to have separate allocation rules for separate expense items (e.g. allocate phone system and information system costs, and Personnel Department costs, according to the number of staff members in each program, allocate malpractice insurance cost by the number of patients served by each program, etc. etc.). In a single program model, like that used for Supported Employment, a single percentage of direct expenses is used to estimate the G&A expenses to add to the program’s total cost.

There is a tabsheet called **“Budget”**, wherein the user can enter, for each individual program in the model, the staffing pattern and the estimated cost of various fixed expense items as well as large initial start-up items specific to that program (e.g. share of facility space). Using “master lookup” tables of all staff names, or titles, and their organization-wide salaries and fringe benefit cost, specific salaries and fringe benefit expenses can be called up and inserted into the proper cells alongside the name of each staff member.

For each program in the model there is a separate tab sheet with that program name on it. Then, within each service program, based on the number of likely users, assigned to it through the “triage” or “assignment” tables, the model will ask for estimates of the units per user as well as the cost per unit if that is how the service will be paid. For programs owned and operated by the organization, the total direct expenses are read into the tabsheet as well as any variable cost items that depend upon users and units. In some cases, the model is designed to calculate the cost per unit by dividing number of units into the program's total cost (including allocated General and Administrative expenses).

In Summary, a typical (generic) stochastic model of a behavioral healthcare system, including EBPs, would likely contain the following attributes or structural features.

- Stratification of population cohorts by attributes that are correlated with their probability of becoming users (i.e., SSI versus AFDC; adults versus children, etc.).
- Stratification of users or clients by cohorts that reflect their potential for differential use of programs and differential service use patterns (i.e., persons with serious mental illness versus people with substance abuse disorders versus those with dual diagnoses versus those with developmental disabilities).
- For each population cohort, the estimated prevalence of each client cohort.
- For each client cohort, the estimated rates of demand for services (i.e., the treated prevalence rates).
- Organizational service capacity, including assumptions about available service time and staff productivity relative to payroll hours, full salary and fringe benefit costs.
- Triage or projected assignment of clients in each cohort to available clinical programs.
- Fixed and variable costs for each direct service program, as well as organizational overhead costs properly allocated to each direct service program.
- Projected distributions of service utilization within each program for each cohort.
- Based on the variable inputs, formulae to calculate for each cohort, within each program, and for all programs combined:
    - Total Cost
    - Costs per unit of service
    - Episode Costs (cost per person for an episode of care within a program)
    - Case Costs (cost per person across multiple programs)
    - PMPM Costs

An Illustrative Stochastic Model of Supported Employment
To illustrate the basic principles of stochastic modeling, we have designed and built a <u>simplified</u> model of a Supported Employment (SE) Program. While this model will not demonstrate all the complexities of a real model, it has a sufficient number of features that make it useful as a didactic tool.

Figure 4 is an illustration of an input sheet wherein the user is allowed to enter values in all the cells with a yellow background. This particular sheet has a feature called simtables. Using this feature the model will be set up to go through

3 simulations, and within each simulation, go through a number of iterations, as specified by the user (e.g. 500, 1,000, 5,000, etc.). On the first simulation of 1,000 iterations the model will use the first value in the left cell of the simtables. On the second iteration the model will use the value in the second cell of the table. In the final simulation it will use the value in the far right cell. This particular sheet is asking for the size of the total population in the geographical area to be served by the SE program. Based on estimated prevalence rate, the model calculates the number of persons with Serious Mental Illness (SMI) and persons who have a Serious and Persistent Mental Illness (SPMI) in the community. From that number the model asks for an estimate of the percentage that would be appropriate to participate in an SE program (e.g. what percentage of all SMI/SPMI persons are adults and non-elderly who might attend such a program. From this number the model is asking the planner to enter an estimate of the percentage of the appropriate persons that will actually be recruited to come into the program. The final calculation is an estimate of the number of persons to be served at some time throughout the fiscal year.

Population and Prevalence

Like most models built for planning and budgeting, this model is based on a one-year fiscal period, although models can be designed for as many time periods as desired. This model is also designed as a “Demand-Need” model. In other words, it is built as if the State is interested in knowing how much it would cost to fund an SE program with sufficient capacity to serve the likely level of demand for a given population. An alternative approach would be to start with a dollar value and model how many persons could be served in a year with that amount of money.

Figure 4: Input Sheet asking for Population and Prevalence estimates

Enter estimates in the yellow cells based on a 12-month fiscal period; July 1- June 30

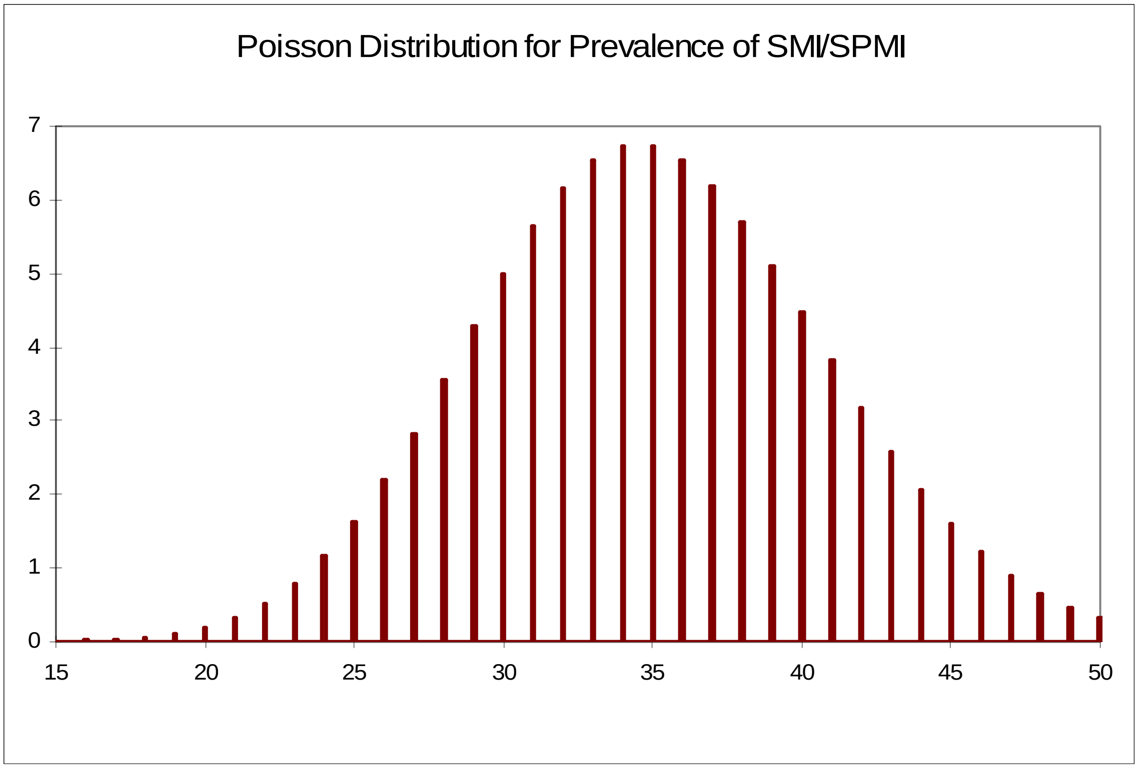
Population and Prevalence Variables		
Number of Total Lives in Community	100,000	All Lives in the General Population
SMI/SPMI Annual Prevalence Per 1000 Lives	35	Rate of Treated Prevalence
Potential Clients/Consumers)	3,500	
Potential % of Clients/Consumers in SE	50%	Rough guess based on age profile of SMI/SPMI population
Potential Number of Clients/Consumers in SE	1,750	
% of Clients Actualy Enrolled in SE	15%	
Number of Clients in SE	263	

Simtables

Urban	Mixed Suburbs-Rural	Rural
100,000	75,000	50,000
35.00	32.00	30.00
50%	40%	30%
15%	12%	10%

While it may appear somewhat arbitrary to enter an estimated annual prevalence of 35, the model, in fact, is using a distribution. Figure 5 illustrates the distribution that is being used.

Figure 5: Distribution Used to Estimate the Possible Values of Annual Prevalence



During each iteration, the modeling software will choose a possible value from this distribution and use that value when calculating all the other formulae in the spreadsheet that depend on this value, such as the calculations of potential clients and actual clients. The sampling will be done proportional to the probability of a given value. Values that are more likely (i.e. the highest part of the distribution around a value of 2) will be sampled more often. Values beyond 5.2 have a likelihood of occurring less than 5% of the time and therefore, in a simulation of 1,000 iterations, a value of 5 or more would be sampled only around 50 times. Note that the Poisson distribution has only integer values.

The planner may choose to enter different values or use different distributions and can do so readily. If another person believes that recent de-institutionalization effortshave modified the estimated annual prevalence rate of 35 per 1,000 lives in the community to40 or 45, it is very easy to enter these alternative values and run another simulation to see what effects a higher prevalence rate would have on total costs. The result of each iteration can be saved as a separate file to allow further comparisons among alternative scenarios. The use of simtables in this model, allow results to be shown as separate distributions of possible costs for each of the three types of geographical areas; urban, suburban-rural, and rural.

Note throughout the model, the simtables allow the planner to enter different average values for each of these three settings, not because they are uncertain variables but because the variation is known to occur and is not readily controllable by the State funding source.

Capacity and Utilization
Given the estimated number of actual clients to use the SE program, the next section of the model attempts to get at some parameters of capacity and utilization as illustrated in Figure 6.

Figure 6: Estimates of Program Operational Time and Time Spent in Program

Program Operations		
Weeks Per Year Program is in Operation	49.0	
Number of Unique Client/Consumers Served Annually	263	From Population Tabsheet

Length of Time in Supported Employment		
Minimum in Weeks	4.0	
Most Likely in weeks	24.0	
Maximum in weeks (Regardless of Fiscal Year Dates)	104.0	
Distribution of Average Weeks per Active Client during the Fiscal Year, in Supported Employment	31.1	Adjusted to Avg Weeks Per Fiscal Year per Client

Weeks of Active Service		
Total Number of Service Weeks (Note 1:)	8,172	Unique clients time average weeks in fiscal
Full Year Equivalent Clients	167	
Ratio of Unique Clients Served to FTE Clients	1.6	Latimer's Range: Min=1.4; Max=3.1; Avg.=1.8
Note 1: Service Weeks are the number of clients/consumers times the Average Number of Weeks per Client per Year		

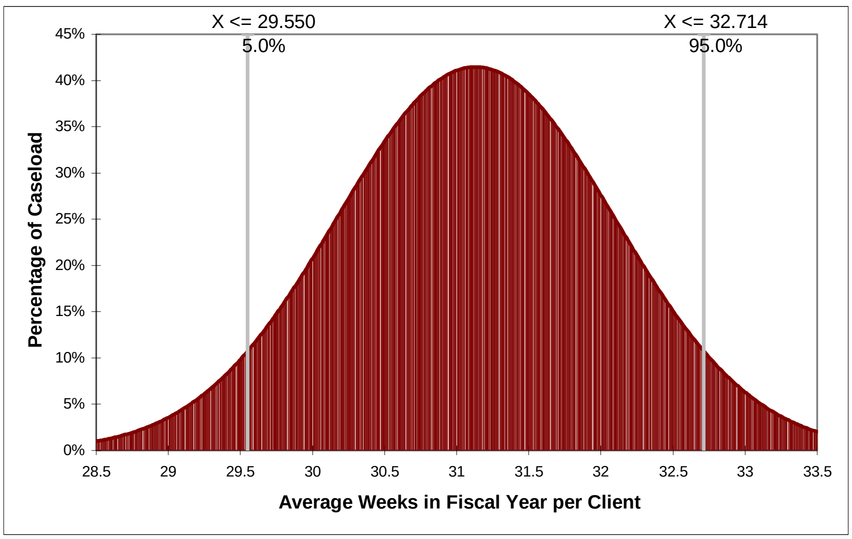
The average weeks the program is operational is based on a full year less the weeks the program shuts down for staff development training, and holidays. A planner may model a program that was operational for 52 weeks by simply changing the entry to 52 weeks.

The three yellow cells ask for the minimum, maximum, and most likely time period the average client will stay in the program. These values are the authors' estimations due to the paucity of published data on the average length of time consumers are active in SE programs.

The estimate of *minimum*, *maximum*, and *most likely* are designed to create a triangular distribution to which further statistical adjustments are made to allow for the fact that the length of time some clients spend in the program overlaps the beginning or ending dates of a fiscal year. Therefore, the average length of time in the program from a client's experience is not necessarily the number of weeks that are of concern to the cost accountant, who is only interested in the average number of weeks spent in the program that occur within the fiscal year. The brevity of this document does not allow for technical adjustments necessary to go from "client's average length of time in program" to "average weeks per client in fiscal year". Latimer made a similar adjustment when he calculated the average FTE clients served within the year, by adding the weeks or months of clients with partial fiscal year attendance to equal a full-time (i.e. full year) equivalent client.

The input distribution in Figure 7 below describes the average weeks in a fiscal year per client.

Figure 7: Distribution of the Average Weeks in Fiscal Year per Client



Statistical Values	
Minimum	1
Maximum	49
Mean	31.13209
Mode	31.13209
Median	31.13209
Std. Dev	0.96198
Variance	0.92541
Left X	29.55
Left P	5.00%
Right X	32.714
Right P	95.00%
Diff. X	3.1646
Diff. P	90.00%

Although the visual display is cut off at the minimum value or 28.5 and the maximum value of 33.5, the actual lower and upper limits, 1 and 49 respectively, are displayed in the narrow table that summarizes various statistical values about the distribution.

Based on the average fiscal year weeks per client, and the number of unique clients, the model calculates the number of service weeks (i.e. one client in the program for one week is one service week).

Staffing Capacity and Variable Client Costs

Once we have an estimate of the demand, or clients that might enter the program, we can use the EBP standard of 18 clients per caseload per Employment Specialist (ES) to estimate the number of Employment Specialist that would need to be recruited, hired and trained. This number can be adjusted as the SAMHSA EBP for SE allows for up to 25 per caseload within the fidelity standards . The percent of direct client staff time generally decreases a few months post placement allowing for increased caseloads. Figure 8 demonstrates needed staff.

Figure 8: Employment Specialist and Support Staff Needing to be Hired

Full Time Employment Specialists Required		
Average Caseload per Employment Specialist (ES)	18	Program Standard
Number of Employment Specialists (ESs) Needed	9.0	Rounded Down
Program Manager / Supervisor	1.0	
Clerical /Receptionist/Van Driver	1.0	

Van Drivers Needed		
Urban	Mixed Suburbs-Rural	Rural
0	1	1

In addition to the Employment Specialist, a significant barrier to the successful operation of a SE Program is the availability of transportation, especially for clients in suburban and rural communities where public transportation is lacking. Therefore, the simtable allows us to vary the number of van drivers to be hired in addition to at least one clerical receptionist. We have estimated one “Program Manager/Supervisor as adequate for a staff of 9 Employment Specialists.

Knowing the core staff size, the model proceeds to estimate costs based on average salary levels in urban, suburban, rural areas. See Figure 9.

Figure 9: Cost Estimates for Staffing and Variable Client Expenses

Annual Costs			
Average Salary for an Employment Specialist	$35,000		
Averaged Salary for Program Manager and Supervisor	$45,000		
Average Salary for Clerical and Support	$18,000		
Fringe Benfits Rate	30%		Depends on owner/ operator and location
Total Salary and Fringe Benefits Expenses		$491,400	
Initial and On-going Training Costs per Employment Specialist	$1,500		
Total Initial and On-going Training Expenses		$13,500	
Other Fixed Annual Operating Expenses (including depreciation)		$83,683	From Start-Up and Annual Budget Sheet
TOTAL FIXED EXPENSES		**$588,583**	
Variable Transportation Expenses per Day per Client	$2.00		
Total Client/Consumer Transportation Expenses		$38,941	Client travels 75% of annual active days
Other Variable Expenses per Unique Client/Consumer (e.g. clothing, grooming)	$100.00		
Total "Other" Variable Costs per Client/Consumer		$26,250	
TOTAL VARIABLE COSTS		**$65,191**	
TOTAL ANNUAL FIXED AND VARIABLE OPERATING EXPENSES		**$653,774**	
ADMINISTRATIVE RATE (Admin. Support)	28%		
TOTAL COSTS, INCLUDING ADMINISTRATIVE EXPENSES		**$836,832**	

Urban	Mixed Suburbs-Rural	Rural
$ 35,000	$ 31,000	$ 28,000
$ 45,000	$ 36,000	$ 32,000
$ 18,000	$ 17,000	$ 16,000
30%	21%	12%
Off-Site Training at Model SE Prg	Coaching plus conferences	On-site Training
$ 4,500	$ 3,000	$ 1,500

Transportation Cost per Day per Case		
Urban	Mixed Suburbs-Rural	Rural
$2.00	$4.00	$7.00
$ 100	$ 75	$ 50

Freestanding	Add-On to Small System	CMHC Sponsored
28%	20%	12%

The Fringe Benefits costs were also roughly estimated but one would expect even more variation within urban-suburban-rural areas based upon the type of organization operating the program. In State-operated programs, fringe benefits may be higher than in those programs operated by private-not-for-profit organizations. These costs will also vary based on the extent to which staffs are unionized. Latimer reported an average Total Cost for an Employment Specialist at $55,668, but it was not clear if that included fringe benefits.

This model estimates training costs at three different levels, but not along the urban-rural dimension. In a customized model the training costs would be known and a simtable would not be necessary. However, for didactic purposes the authors envisioned sending staff to a mature model off-site SE program where they would get on-the-job coaching and learn from a successful SE staff. Another scenario would be sending staff to various conferences to learn about SE operations. The least expensive option would be to have on-site training and coaching of staff. The other fixed cost items on this sheet are based on calculations on the budget sheet that has not yet been presented and discussed.

The variable cost items are based on the number of clients coming into and through the program. The most significant client related cost is transportation, which also comprised a separate line in Latimer's cost summary. For clients in urban areas we assumed $2.00 to cover daily costs of public transportation, and $4.00, and $7.00 in suburban areas and rural areas, assuming van transportation averaging 20 miles per client per day and assuming a per mileage cost of $0.35 for operating a van (an item that was expensed in the start-up budget for suburban and rural areas) The van driver was itemized separately in the above staffing pattern. The final variable cost was for the expenses related to getting the consumer “job-ready”, (e.g. appropriate clothing, personal grooming expenses, alarm clocks, etc.).

Finally, an administrative cost rate of 28% was applied to the total of fixed and variable costs to arrive at total costs. The 28% administrative rate is also likely to vary based on settings. If the SE program stands alone, it must have separate administrative positions and expenses; whereas, if it was part of a larger organization with multiple programs, its administrative overhead might be as small as 12%. While it is not entirely clear in Latimer’s article how his 7 responding sites performed cost accounting and allocations, their reported “Overhead” costs ranged from a low of 10.5% to a high of 53.9% with an average of 26.3%.

Start-up and Annual Operating Expense

The next section of the model is the start-up budget and the estimated annual operating expenses (other than staff salaries and fringe benefits which were already calculated). Figure 10 presents an illustrative Start-up budget and an Annual Operating Budget. In many circumstances start-up cost may be forwarded to the vendor as a separate grant or contract while annual costs are paid for in some sort of fee-for-service or case rate payment method. In this model certain start-up cost, regardless of how they were paid for are amortized on a depreciation schedule suited to the type of item.

Figure 10: Start-up Costs and Annual Operating Expenses

Start-Up Cost other than Staff and Training (Start-up to be paid to Program Vendor in separate Contract) and Annual Operating Expenses, including depreciation, but excluding Staffing (see Model)						Start-Up Cost	Depreciation Yrs	Annual Expenses & Depreciation
Facility Deposit						$5,000		
Facility Rental/Lease	500	Sq.Ft per ES Team	500	Common Space and Admin				
	$ 18	Cost per Square Ft (including uitilities and maintenance)						$18,000
Vehicle for Transporting Clients (See Note 1.						$0	5	$0
Desks, Chairs & File Cabinets for staff & consumer use	$600	per Staff	10	For Consumers	3	$7,800	5	$1,560
Office Supplies	$500	per month						$6,000
Computers for staff and consumer use	$1,000	per Staff	10	Consumers per day	3	$13,000	5	$2,600
Software	$800	per PC				$10,400	5	$2,080
Xerox Machine						$5,000	5	$1,000
Fax Machine						$500	5	$100
Phone system						$21,760	10	$2,176
Cell Phones and Annual Service Plan	$200	per cell phone	$200	Monthly Srv Plan		$2,000	3	$24,667
Recruitment						$10,000	20	$500
Accreditation - CARF						$10,000	20	$500
Insurance								$8,500
Printing & Publications								$6,000
Advertising and Job Marketing								$10,000
Total						$85,460		$83,683

Urban	Mixed Suburbs-Rural	Rural
$18.00	$16.00	$14.00
$0	$12,000	$12,000

Notes:

Note 1:Based on Geographical Context. Annual cost based on $.35 per mile driven - See Variable Costs in Model Tabsheet

Notice that within the urban scenario there are no expenses budgeted for a van. Recall that all yellow cells can be changed if another planner wishes to use different cost estimates.

Outcomes

The next section of the model is an attempt to highlight the importance of including outcome measures in cost models- the only way to truly judge the cost-effectiveness of programs. Unfortunately, there is little reported on the outcomes achieved by high-fidelity ES programs. Many states utilize outcome measures related to number of successful placements, number of successful months on the job, salary at minimum wage or better and whether or not the client receives fringe benefits.

While the effectiveness rate presented here is totally hypothetical, and is unlikely to be achieved in the first year of an ES program's operation; especially since it takes 6 months of job stability to be counted as a success and in the first year clients are unlikely to get to the point of successful program completion by the sixth month of operations, we include the variable to illustrate how one can determine cost per successful outcome. Figure 11 presents the small illustration of estimating outcomes. In reality, many EBPs have more complex measures but data is still lacking.

Figure 11: Estimation of Successful Outcomes

Program Performance		
Percentage of Client/Consumers Experiencing Successful Placement (See Note 2)		58%
Note 2: Successful Placement is at least 6 months on a job paying at least minimum wage.		
Number of Successful Placements		153

Summary of Total Cost

The final section of the Model is a summary of total costs and various cost indices. Total Direct Costs are distinguished from Total Cost (including administrative costs) in order to compare "average" urban results with the data collected from the 7 SE sites surveyed by Latimer. This model represents the average values for only the urban scenario because the software is designed to use the first, left-most value in each simtable, which is consistently the most expensive option. Below, the model presents and compares the distribution of possible values that result from doing three consecutive simulations, one for each level of the simtables. Figure 12 presents the Cost Summary information.

Figure 12, Summary of Total Costs and Cost Indicators

Cost Summary		Latimer's Data ** Min	Max	Average
Direct Cost (Expenses before Administrative Costs are applied)	**$654,590**	$147,039	$409,443	$251,272
Direct Cost per Unique Client/Consumer Served Within the Year	$2,487	$860	$2,723	$1,361
Direct Cost per FTE Client	$3,915	$1,423	$6,793	$2,449
Direct Cost per Employee Specialist	$72,732	$37,339	$49,013	$44,082
Total Costs with Adminstrative Costs Included	**$837,876**	$198,581	$472,554	$317,310
Total Cost per Unique Client/Consumer Served Within the Year	$3,184	$993	$3,677	$1,644
Total Cost per FTE Client	$5,011	$1,602	$5,036	$2,896
Total Cost per Employee Specialist	$93,097	$52,506	$66,194	$55,668
Total Cost per Client Service Week	$102.26	NA	NA	NA
Total Cost per Successful Job Placement	$5,489	NA	NA	NA

** Latimer's Averages from Page 404 of manuscript are based on Direct Costs because he says in a later paragraph that overhead would increse these "15 to 30 percent". Latimer's data is based on the value of the 2001 dollar.

The current model's Direct Cost and Total Cost are much higher than the costs reported in Latimer's article. However, Latimer's article is not specific enough to demonstrate how each of the reporting sites recorded and reported either their direct or overhead expenses. For example, this model includes all start-up cost amortized over the fiscal time period. According to a footnote in Latimer's article, similar start-up costs for a few of these sites were covered by federal grants and were not included in the reported costs.

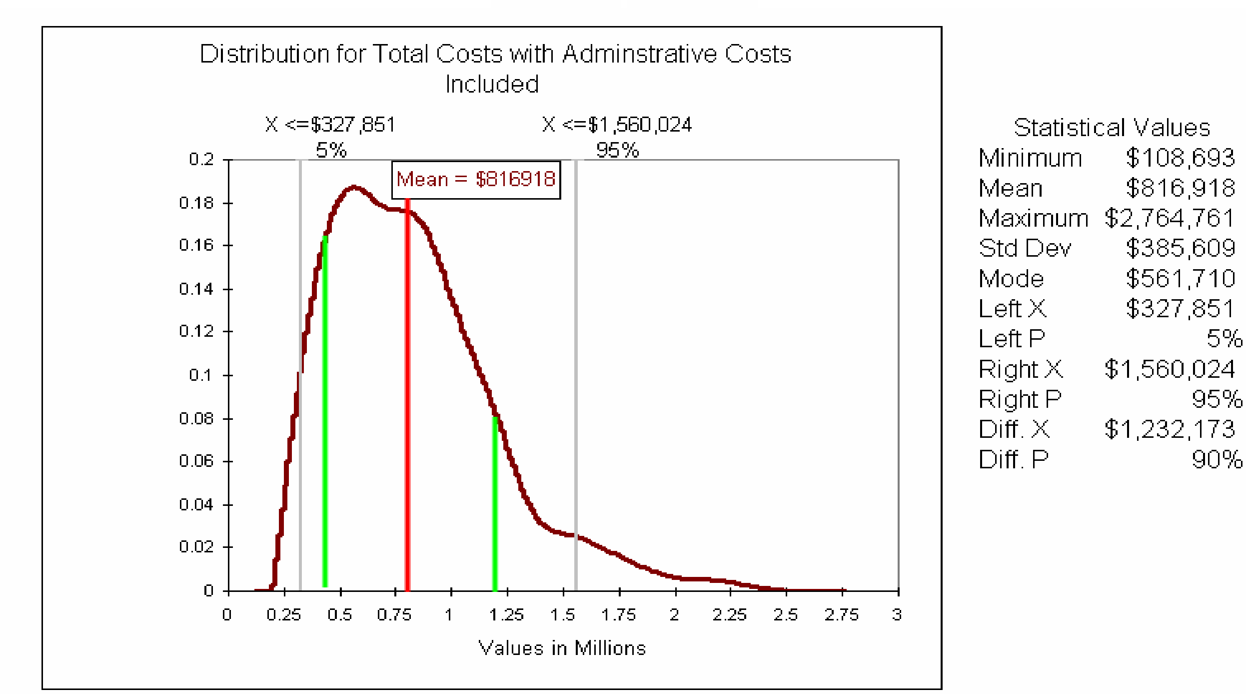
Results of Three Simulations

A simulation was run for each of the three scenarios. There were 1,000 iterations per simulation, meaning that a potential value was sampled from every distribution in the model, allowing the software to collect and graph 1,000 possible values for each “output” variable the authors selected. While every variable within the Cost Summary table could have been selected, in the interest of brevity, only the results and interpretation for the output variable of Total Cost (including administrative costs)is presented.

Total Costs, including Administrative Costs

It is important to recall that this model is based on a one year fiscal period and designed to estimate costs for a given population, which varies with each of the three scenarios, Urban, 100,00; Mixed Suburban-Rural, 75,000; and Rural, 50,000. Figure 13 represents the distribution of possible costs for the Urban scenario.

Figure 13: Distribution of Total Cost (including administration)



The distribution in Figure 13 represents 100% of all likely values based on 1,000 iterations and 1,000 combinations of sampled values from the various input variables. In Figure 13 the mean, with a value of $816,918 is designated by a red vertical line. The green line marks values that are within one standard deviation from the mean in both directions. The area of the graph within these two green lines, $431,309 to $1,202,528, represent 67% of the total area of the graph. Therefore, the most likely cost for a urban-based program with these operational parameters is between $431,309 on the low side and $1.202,528 on the high side.

Unlike a symmetrical normal distribution, this distribution is “skewed” to the right, meaning that there is a small probability of larger values, like $2.25 million. The likelihood of such high values has the effect of “dragging up” the mean, such that it is higher than the median value of $800,000, and the modal value (the value with the highest frequency of occurrence) which is $650,000.

As discussed in the earlier sections of this manuscript, the mean of a skewed distribution may distort the planner’s estimation of what is likely to happen. Since many phenomena in health and human services are distributed this way (e.g. length of stay, cost per episode, cost per case; Pareto’s 80/20 law – 20% of the patients consume 80% of the resources, etc.) planners must take into account the fact the mean values alone is not useful without a more complete picture of the entire distribution of likely outcomes.

Based on this distribution one could make probability statements relative to any Total Cost value. For example, the model, based on all the assumptions that were made, estimates an average Total Cost of $816,918. Ninety-five percent (95%) of all the possible values for Total Cost lie between a low of $327,851 and a high of $1,560,024. While this wide range may not be comforting, it is probably realistic. How often has anyone seen a new program come in exactly at the budget costs?

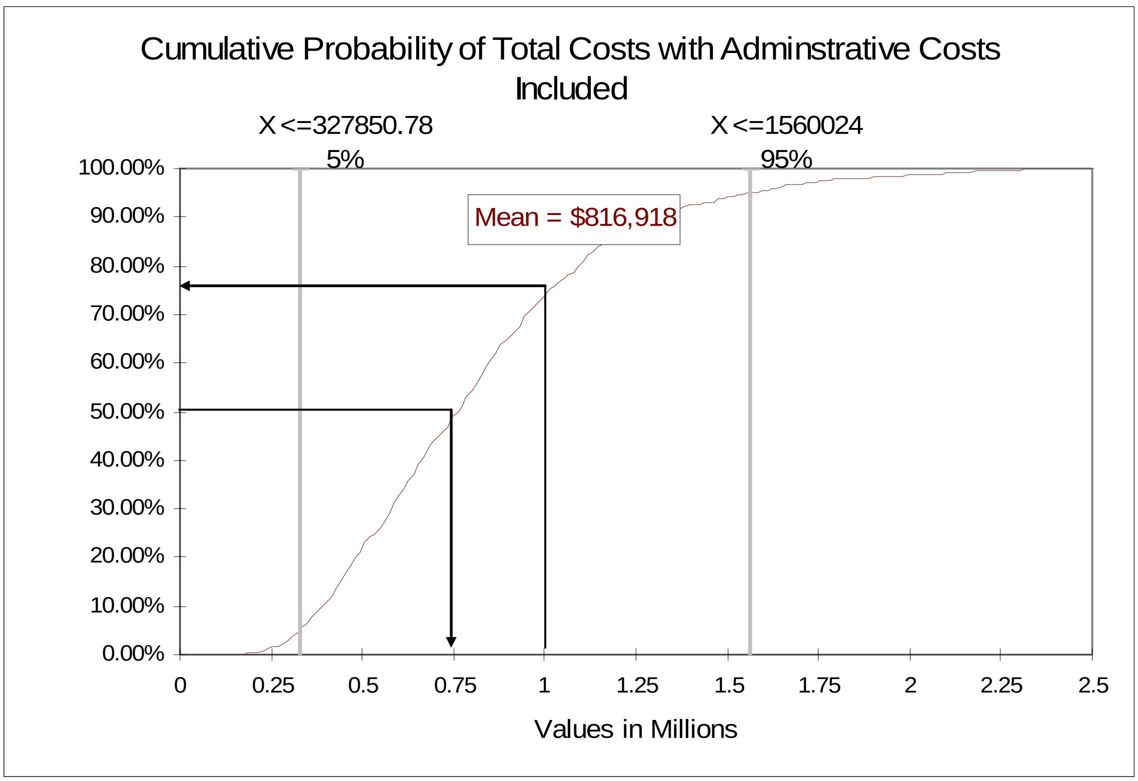
The model allows the planner to look realistically and explicitly at the variables driving the possible cost variation. In the real world, the planner can begin to think about what operational steps need to be taken to reduce the variability and the uncertainty in knowing what to budget; (i.e. to get a “tighter” estimate). One can do little about possible prevalence rates, given the many unknown and uncontrollable factors that drive prevalence. Yet this model, which is designed to be sensitive to prevalence by staffing to meet demand, is no doubt primarily driven by the prevalence factor. Perhaps the planner would want to consider looking more closely at the epidemiological data for their State in order to justify using a “tighter”, less variable, distribution to describe the likely treated prevalence.

Since almost 11% of total direct costs are variable costs, the planner may want to consider how the variability in clients’ average time in program is driving total cost variability.

As the planner does her research she may find there are consumer attributes that correlate with successful placement, and redesign the model to accommodate a screening process. This additional process would add cost but possibly reduce variable costs as a result of screening out consumers that are not likely to fully benefit. Furthermore, this step would increase the number of successful placements relative to the total costs, thereby reducing the most important index, cost per successful placement.

A frequency distribution such as Figure 13 shows the relative probability of all possible values along the horizontal, “X-Axis”. Another way of looking at the same information is to convert it to a cumulative probability distribution, often done when one is negotiating with another party to accept responsibility (financial risk) for operating a given level of program on a fixed budget. Figure 14 illustrates exactly the same data as does Figure 13 but in a slightly different format. In Figure 13 the vertical, “Y-Axis” was the exact probability of a specific value along the X-Axis. In Figure 14 the Y-Axis represents the cumulative probability of any given value or less on the X-Axis.

Figure 14: Cumulative Probability of the Total Cost Reaching or Exceeding a Given Value

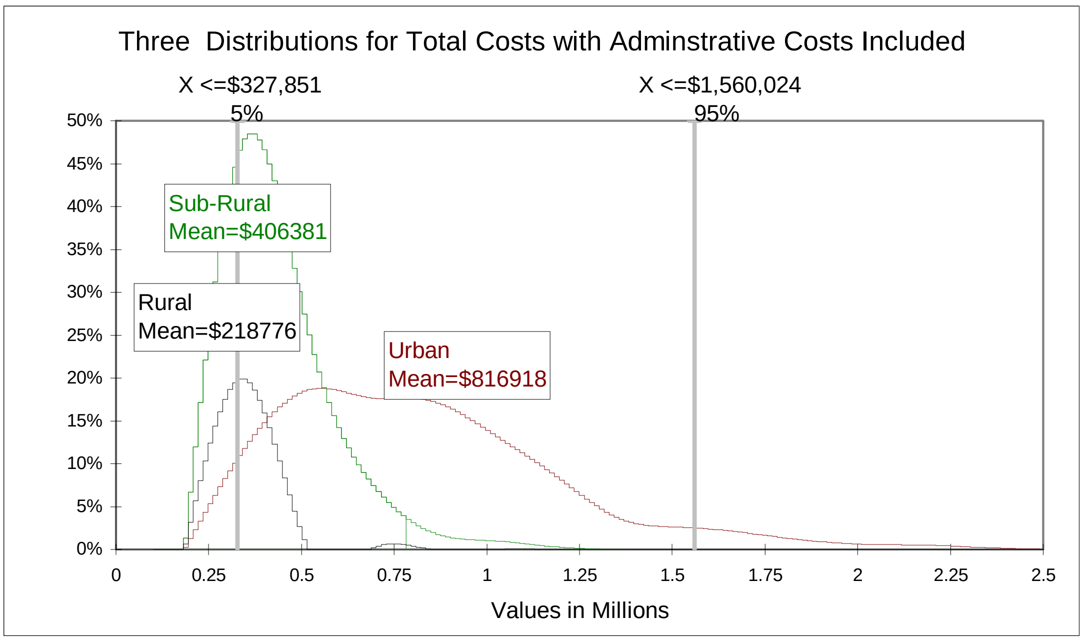


The way to interpret Figure 14 is as follows: the cumulative probability on the Y-Axis ranges between 0 and 100%. The probability of each X value (cost) is added to the probability of the next X (cost) value. Therefore, the curve represents the cumulative probability of any given cost or a cost that less than X. For example, if you go to the 50^{th} percentile on the Y-Axis and draw a line over to the cumulative curve and trace the line down to the X-Axis, you could say that 50% of all possible costs values will be at $750,000 or less. Notice that because the distribution in Figure 13 is skewed to the right, the average value is greater than the median (50^{th} percentile) value. Or you could move from any value on the X-axis (e.g. 1 million) and draw a line up to the curve and over to the Y-Axis, to identify the probability that the Total Cost will be no greater than that X value of

$1 million. In the above figure you could say that the probability the costs will be $1 million or less is 78%. Therefore, there is a 22% (100% - 78% = 22%) chance it will exceed $1 million. So given a contract for $1,000,000, the vendor using this model, and operating within the ranges of the various factors within the model has a 78% chance that the Total Costs not exceeding that amount of revenue.

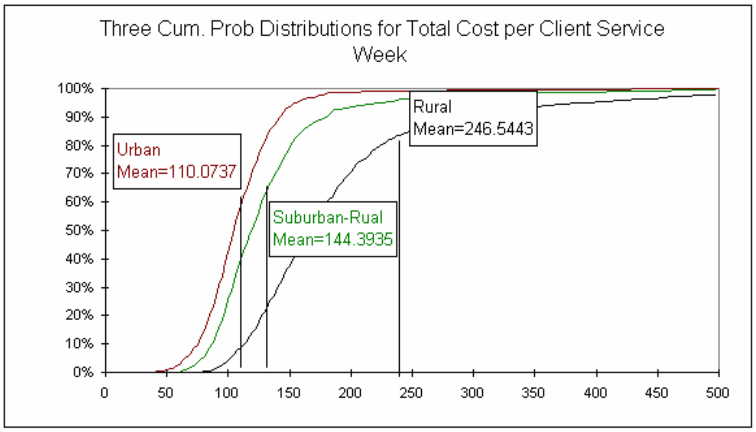
We can also compare the distribution of costs for all three scenarios. This information is presented in Figure 15.

Figure 15: Comparisons of Total Costs for Three Scenarios



The urban scenario is the red distribution, the suburban-rural scenario is the lighter green distribution, and the rural distribution is the black distribution. As expected, the rural scenario is not only less costly, with an average of $218,776, it has far less variability or greater certainty. The lower distribution of costs reflects the lower values assumed for rural areas in each "simtable". The lower variability is because we assumed a lower prevalence rate, and used a Poisson distribution to describe the likelihood of variation in the rate. As the mean of a Poisson distribution is reduced, so is it variance or variability. Because we assumed a higher prevalence rate for urban areas, the variability of that rate also went up, simulating less certainty about the number of consumers who will enter the SE program. We also assumed greater variability in urban salaries and related expenses. Since the number of consumers drives staffing costs and all variable costs, it is not surprising that there would be more uncertainty in budgeting for a SE Program in an urban area

Figure 16: Comparison of Three Scenarios in Average Cost per Service Week



	Urban	Sub-Rural	Rural
Minimum	$33	$54	$67
Mean	$110	$144	$247
Maximum	$1,186	$12,199	$48,822
Std Dev	$54	$388	$1,545
Variance	$2,931	$150,441	$2,388,386
Mode	$122	$108	$156
Left X	$65	$65	$65
Left P	5.0%	0.8%	0.0%
Right X	$156	$156	$156
Right P	95.0%	83.2%	43.2%
Diff. X	$91	$91	$91
Diff. P	90.0%	82.4%	43.2%
5th Perc.	$65	$79	$102
95th Perc.	$156	$226	$388

In this particular simulation both Urban and Suburban-Rural had very similar Average Costs per Service Week, with Urban being only slightly lower. The rural average was the highest, because it had fewer clients to divide into the same amount of fixed costs as the other two scenarios. Clearly, in funding a rural SE program with potentially few consumers, it would be important to consider every conceivable way of making the costs more variable than fixed, perhaps at the expense of maintaining certain fidelity standards. Perhaps consumers could be served initially in their home. Perhaps the team of ESs could occupy a smaller facility, and use mobile information technology (laptops and mobile phones) to maintain contacts with employers and consumers. It would be worth exploring Latimer's data in more detail to determine if rural-based programs indeed had higher average cost per FTE client (a correlate of service weeks).

Benefits of Stochastic Modeling

Analysts can use available software, such as Crystal Ball or @Risk, as add-ons to Excel to create spreadsheets in which all the possible variable inputs (e.g., average units per user, average cost per unit) can be entered as distributions rather than fixed, "average" values. The regular formulae available through Excel can then be used to link these input variables to the calculations of all the desired outputs (e.g., cost per member per month). All the "downstream" calculations (e.g. Total Costs) that depend on these variable inputs would also change each time the inputs change. If the analyst were to run 1,000 iterations, 1,000 possible "results" would be generated. Using these software add-ons to Excel, the iterations can be run quickly, and the relative probability of alternative output values could be plotted. Key performance measures like program cost, total costs, as well as cost per client/consumer, and cost per positive outcome; are re-iterated thousands of times; each time sampling likely possible values from all the various distributions in the model. The software samples probable values proportional to the likelihood of their occurrence. With each iteration it carries out all the calculations based on all the formulae written into the model, and produces an answer for the value of the key output variables that the user want to know.

If an organization runs a thousand simulation trials, they'll get a thousand estimates of the possible values of their key output variables (e.g. total cost, cost per unit, and/or cost per patient user). They will also see that there is an average expected value as well as a degree of variation above and below that average value. In this way, organizations can know, for any given output they choose, what the probabilities are for each possible value. If an organization wants to be, say 70% certain, that they will operate at or below a given cost in a particular program, the model could tell them the probability of operating at or below that cost given all the input assumptions and values they have entered.

A payer organization, like State government, may be interested in such a model to establish a “fair” unit rate, or a “fair” case rate, or an adequate total aggregate grant to give to an EBP program in order for it to succeed with a given level of certainty. “If we give this organization this specific amount of funding, it has a 60% chance of financial solvency and success of operating within the fidelity standards that have been established in the contract”. Other State’s may feel an amount that is “fair” is equivalent to a 50% chance of financial solvency operating with the performance standards. The model could also be used to estimate a fair share of costs when revenue sources are shared among payers.

If the model was built to reflect the operation of a multi-program system of care, once populated and calibrated to reflect actual performance of the “real-world” system, the users could use it to test alternative strategies for cost savings or reduction in the uncertainty of each year’s programmatic cost and performance. One strategy of system redesign would be to redirect patients. If an organization has 12 different service programs, including high-cost inpatient care they might try the strategy of creating one or more EBPs as new programs in the system and adding it to the model in order to reduce the number of clients needing to use inpatient or to reduce the length of inpatient stays. Another strategy could involve increasing or decreasing the funding available, or changing the expected number or cost of units or utilization patterns. If, for example, the intensity of service is increased, so that a person could get through a program in 26 rather than 36 weeks, the estimate for processing clients through a system could be increased as well.

A well-designed stochastic model also allows organizations to look at fixed cost as well as variable costs. For the payer, it is important to have as few fixed costs as possible. If there are large fixed costs, it is important that programs are fully utilized. The model can be designed to compare an organizations’ actual utilization against its capacity for utilization. It can indicate which programs are efficiently operated or under-utilized. Some organizations have looked at this data and come to the conclusion that operating a program makes less sense than purchasing it from other organizations who are willing to sell it to them on a per unit basis or a per client basis at a lesser cost than they can do it themselves.

What makes stochastic modeling and simulation different than most planning and budgeting systems is its relative non-dependence on historical data. While organizations should look at their historical data, (in the form of frequency distributions) to see how effectively and consistently they have been operating, using historical data to establish future costs is not a sound approach. Historical data is also misleading because it assumes that provider', client' and system' behaviors won't change. In fact, these aspects do change considerably. The model can be used to compare what a new system is doing with their past operations. Three months into a new system, for example, an organization can get actual data that they can put into the model to compare where they are with where they thought they would be when they built the model and projected their future. In this sense, the model becomes a tremendous monitoring, planning and budgeting tool because managers can see where projections were off since the tool expresses mathematically and explicitly all the assumptions about needed capacity, actual assignment of clients to alternative service patterns, client-centered utilization within programs, and actual cost per service unit or cost per client served. Not only are all these goals now explicitly stated, various clinical and financial perspectives are brought together, so that one part of the system can see how their actions affect another part. Formerly independent actions are now interdependently defined. Historical data by itself is totally inappropriate for modeling the utilization, cost and capacity of EBPs. Prospective methods that allow for possible variation and uncertainty is a way of voicing strategies about what the future could be, rather than simply assuming that history repeats itself.

One current trend involves examining client and program cost variations and the risk such highly variable cost drivers engender in managing costs. Systems ought to start thinking of "risk" not as high cost, but as uncertainty. A "high risk client" has a utilization pattern of uncertainty. Often people spend time managing a single high-cost client who is non-compliant rather than investing time in programs to improve compliance and reduce variation. A properly designed model can point out these cost management issues. EBPs are an effort to reduce variability in programs and replace it with factors known to produce consistently good outcomes. Systems that try to plan their futures on the basis of past data are analogous to someone trying to drive a car forward by looking through the rear-view mirror. Table 5 summarizes some of the benefits to be derived from the use of formal stochastic models in the planning and budgeting process.

Table 5: Benefits of a Formal and Comprehensive Stochastic Model

Models make explicit the assumptions underlying the Business Plan/Budget and the operational procedures to be used in the System.

Models support contract negotiations among payers and providers around alternative budgets, case rates, PMPM rates, stop-loss levels, withholds, and bonus systems.

Models can estimate post-contract savings/(deficits) and surplus/(losses) to be realized through alternative use of resources, and use of alternative pathways of treatment for different patient groups.

Formal models create a shared vision among administrative and clinical leaders with respect to the ongoing expectations for program performance.

Models can reinforce a comprehensive view of the entire system of care, one that allows the interdependency and continuity across programs to be explicitly displayed and analyzed.

Models free program planning and financial forecasting from the conceptual constraints of "the average patient" mode of thinking and incorporate a more realistic assumption that "variation happens".

Stochastic models support a philosophy of Continuous Quality Improvement (CQI) that suggests that quality is reflected in the reduction of controllable variation and matching patient variations to individualized treatment.

An initial Estimation Model can be converted into a Tracking and Trending Model that supports an ongoing comparison of the assumptions and predictions made when the program was started with the actual results as they show up in utilization and cost reports.

Models support the planning and analysis of new programs needed in order to better manage total system costs through estimate of the Return on Investment (ROI) in new programs.

Further References and Other Sources of Information

The above discussion details the use of a rigorous computer-based analytical technique to help make better decisions concerning the planning and budgeting of EBPs. Extending this technique to all aspects of planning and budgeting operations would allow a mental health organization to use logical and quantitative techniques to support better decision-making about all aspects of their operations. The key lesson here is that clinical programs designed to help people with mental health problems can be more successful if supported with the analysis suggested by the demonstration model for supportive employment discussed above.

Understanding and quantifying the key cost-driving variables (simplified to be Number of Users X Number of Units of Service X Costs per Unit of Service), in combination with the realization that uncertainty (variability or risk) influences each component of such an analysis, yields a powerful approach to planning and budgeting a particular program or course of action. Although these analytical techniques are based on rigorous decision theory, they are very accessible to mental health professionals who may only have rudimentary training in mathematics and statistics. This section will provide additional sources of information that can be used to make these powerful techniques part of an ongoing approach to planning and funding behavioral health services.

Decision theory is a well-established field that has been applied to a broad spectrum of analytic problems, including behavioral health. While it is based on mathematics, statistics, and behavioral sciences, it remains accessible to most practitioners because its fundamentals are based on easily understood concepts. Mental health professionals interested in gaining a deeper understanding of the theory and practice of decision theory and risk analysis can turn to a rich literature including:

- Clemen, Robert T., (1990), *Making Hard Decisions*. PWS-Kent Publishing Company, Boston.

- Megill, Robert, E., (1984), *An Introduction to Risk Analysis*, PennWell Books, Tulsa, Oklahoma.

- Raiffa, Howard, (1968), *Decision Analysis: Introductory Lectures on Choices under Uncertainty*, Addison-Wesley Publishing Company, Reading, Massachusetts.

- Vose, David, (2000), *Risk Analysis,* 2nd Edition, John Wiley & Sons, West Sussex, England.

In addition, many standard textbooks in the field of corporate finance and economics contain excellent discussions of risk analysis and decision theory.

The theory described in this extensive literature has become accessible to non-mathematicians in the past 10 years due to the innovation spawned by the microcomputer. Standard programs like Excel have offered average users both exceptional computational power and ease-of-use. The addition of specialty Add-in programs such as @RISK (Palisades, Inc) and Crystal Ball, (Decisioneering, Inc.) such popular programs as Excel are transformed into extremely powerful, yet easy-to-use risk analysis platforms.

Pareto Solutions has been working with risk analysis and other computer-based analytical techniques to help improve operational, outcome, and financial efficiency in the fields of both mental health and child welfare. In both fields the objective has been to optimize the services provided for clients under the constraints of limited budgets. The techniques that we have implemented have become extremely valuable management tools to maintain and improve services in a challenging financial climate. These techniques can be successfully and cost-effectively integrated into mental health organizations if management understands and commits to the effort, and if cooperative alliances are formed to share the implementation costs.

For example, our statewide experiences in Michigan, Florida, and Colorado has helped us evolve an approach that can make the techniques described above both affordable and accessible to even local organizations. A key factor is cost and experience sharing. Specifically, a group of mental health organizations (e.g. counties, private providers) with similar financial and operational challenges pool their resources to have a professional consultant such as Pareto Solutions build a "generic" model that each can use. Additional participants could include local, (i.e. county or district level organization) and larger umbrellas groups such as State funding and/or administrative entities. This generic model must successfully encompass all of the relevant issues faced by all the participants. The generic model is built using a common computer platform such as Microsoft Excel (with the additional power of an add-in such as @RISK).

Typically the process would involve several phases. The first phase would be to design the model. Representatives of all participating organizations would meet with the consultants to define the programs and financial structures to be modeled and to define the relevant input and output variables for the model. In phase two, the consultant would build a draft model that was sufficiently detailed to accommodate all of the factors and concerns agreed upon by the group. The resulting model would be generic in the sense that it would be usable by all members of the group to do meaningful analysis, regardless of nay differences in their size or demographics. For example, although one organization may own and operate all services within their system, while another may purchase them, switches in the model allow one to move between budgets and fee schedules as respective methods for entering cost factors. Upon review by the participants, this first drat model may need further modifications or refinements. Phase three would involve training to enable the participants to take over the use of the model (reducing/eliminating the need for further consultant participation). As part of this phase, participants would begin use of the model under the supervision of the consultant, which may lead to even more fine-tuning of the model. Throughout these three phases, periodic meetings of the group and the consultant may lead to group-initiated enhancements (and further cost-sharing). In its final phases some organizations may request optional customization of the model to meet some of their unique needs.

Bibliography

1. Broskowski, A., and Smith, M. (2001) *Estimating the Cost of Preventive Services in Mental Health and Substance Abuse Under Managed Care*. DHHS Pub. No. SMA-02-3617R. Rockville, MD: Center for Mental Health Services, Substance Abuse and Mental Health Services Administration.

2. Dixon L., et. al., (1993) Modifying the PACT Model to Serve Homeless Persons with Severe Mental Illness, *Psychiatric Services*, 46 (7) 684 - 688.

3. Goldman, et al., (2001) Policy implications for implementing evidence based practices, *Psychiatric Services*, 52:1591-1597.

4. IPS Model Fidelity Scale, New Hampshire-Dartmouth Psychiatric Research Center @www.dartmouth.edu.

5. Latimer, E.D., Bush, P.W., Becker, D.R., Drake, R.E., Bond, G.R., (2004) The cost of high-fidelity supported employment programs for people with severe mental illness, *Psychiatric Services,* 55 (4). @www.ps.psychiatryonline.org

6. National Association of State Mental Health Program Directors. (Dec 1996) Position statement of employment and rehabilitation for persons with severe psychiatric disabilities. @www.nasmhpd.org

7. National Governors Association. (Feb 2000) Developing effective publicly funded jobs programs for low-skilled individuals. Contact Neil Ridley, Social Economic and Workforce Programs Division @www.nga.org

8. National Mental Health Association. (2004) Targeted Technical Assistance Project of the National Association of State Mental Health Program Directors and through a Center for Mental Health Services Block Grant. Supported Employment for Persons with Psychiatric Disabilities.

9. Phillips, S.D., Burns, B.J., Edgar, E.R., Mueser, K.T., Linkins, K.W., Rosenheck, R.A, (2001) Moving assertive community treatment into standard practice. *Psychiatric Services*, 52 (6), 771-779.

10. Stein L., and Santos, A., *Assertive Community Treatment of Persons with Severe Mental Illness.* Publishers: W. W. Norton & Company, Inc., 500 Fifth Avenue, New York, NY 1010.

11. Supported Employment Implementation Resource Kit, draft version (2002). @www.mentalhealthpractices.org.

12. Torrey, W. et al., (2001) Implementing evidence based practices for persons with severe mental illnesses. *Psychiatric Services* 52:45-50,

13. Virginia Commonwealth University Rehabilitation and Training Center on Workplace supports through a grant with the National Institute on disability and Rehabilitation Research (2004)@ www.vcu.edu