Data Analysis and Simulation Approach to Capacity Planning

By

Melvis Ngemasong Ngimndoh Chafac

MS Industrial Engineering
Purdue University, 2012

BS Systems Engineering
United States Military Academy, 2010

Submitted to the Institute for Data, Systems and Society
in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE IN ENGINEERING SYSTEMS

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2015

© Massachusetts Institute of Technology 2015. All rights reserved.

Signature redacted

Author.................................................................

Institute of Data, Systems and Society
September 08, 2015

Signature redacted

Certified by ...... .................................................

Gordon Kaufman Professor of Management
Professor of Behavior and Policy Sciences and Engineering Systems

Authorized by...... .................................................

Munther A. Dahleh
William A. Coolidge Professor of Electrical Engineering and Computer Science
Director, Institute for Data, Systems and Society
Data Analysis and Simulation Approach to Capacity Planning

By
Melvis Ngemasong Ngimndoh Chafac

Submitted to the Engineering Systems Division
on September 08, 2015 in Partial Fulfillment of the Requirements for
the Degree of Master of Science in Engineering Systems

Acknowledgements:

I could not have come this far on my own, but thanks to the giants on whose shoulders I stood I am finally here. To all the professors, supervisors, mentors and friends who supported me in one way or another during my time at MIT, thank you. I am especially thankful to my family (and Ms Jackie McGurk). You stood by me through thick and thin and gave me the strength to rise above all hardship. I am most of all thankful to God, for all He has done in my life, and for changing my direction. I feel blessed.
ABSTRACT

Data Analysis and Simulation Approach to Capacity Planning

By

Melvis Ngemasong Ngimndoh Chafac

Submitted to the Engineering Systems Division
on September 08, 2015 in Partial Fulfillment of the Requirements for
the Degree of Master of Science in Engineering Systems

ABSTRACT

In 2012, President Obama signed an Executive Order to improve access to mental health service for active duty members and for veterans. Two years later, in 2014, the President outlined 19 new executive actions to improve the lives of service members with a focus on improving access to mental health care. These actions placed a priority on improving the capacity to provide mental health care. This thesis examines ways of improving the capacity of the mental health system with a focus on system redesign. I review capacity planning, provide a literature review of simulation methods and present a simulation, and data analysis of Site Alpha, a U.S. Army Installation. I also use causal loop diagrams to explore other feasible scenarios that affect care capacity. The key take-away from this work is that system inefficiencies should be dealt with before more resources can be effectively added and used in the system. Another pertinent finding is that the distribution of the providers in the system should be improved. The system also contains high utilizer patients who must be considered when planning for care.

The mental health system is extremely complex and risks becoming even more complex. However, by adopting a holistic, systems approach to capacity planning the complexity can be managed.

Thesis Supervisor: John Carroll

Title: Gordon Kaufman Professor of Management
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>iii</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vi</td>
</tr>
<tr>
<td>CHAPTER I: CAPACITY GAP IN THE MILITARY MENTAL HEALTH SYSTEM</td>
<td>1</td>
</tr>
<tr>
<td>CHAPTER II: PERFORMANCE MEASUREMENT, QUALITY AND CAPACITY PLANNING</td>
<td>13</td>
</tr>
<tr>
<td>CHAPTER III: HEALTH CARE SIMULATION MODELS &amp; MODEL CONCEPTUALIZATION</td>
<td>27</td>
</tr>
<tr>
<td>CHAPTER IV: DATA ANALYSIS AND SIMULATION OF THE INTENSIVE OUTPATIENT UNIT AT SITE ALPHA, 2009-2010</td>
<td>40</td>
</tr>
<tr>
<td>CHAPTER V: INSIGHTS, POLICY IMPLICATIONS AND CONCLUSION</td>
<td>57</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>72</td>
</tr>
<tr>
<td>Appendix A</td>
<td>99</td>
</tr>
<tr>
<td>Appendix B</td>
<td>107</td>
</tr>
<tr>
<td>Appendix C</td>
<td>109</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table A: System design and the Quadruple Aim</td>
<td>24</td>
</tr>
<tr>
<td>Table B: Different types of models used in the health care literature</td>
<td>31</td>
</tr>
<tr>
<td>Table C: Average daily captured demand for each month</td>
<td>42</td>
</tr>
<tr>
<td>Table D: Service times per clinic</td>
<td>46</td>
</tr>
<tr>
<td>Table E: Number of providers by clinic</td>
<td>48</td>
</tr>
<tr>
<td>Table F: Simulation outputs</td>
<td>51</td>
</tr>
<tr>
<td>Table G: Model outputs with the addition of 1 psychiatrist</td>
<td>53</td>
</tr>
<tr>
<td>Table H: Summary of capacity planning methods</td>
<td>102</td>
</tr>
<tr>
<td>Table I: Comparing different simulation modeling techniques</td>
<td>103</td>
</tr>
<tr>
<td>Table J: Example CPT codes – used to estimate service times</td>
<td>105</td>
</tr>
<tr>
<td>Table K: Example MEPRS3 Codes</td>
<td>106</td>
</tr>
<tr>
<td>Table L: Transition probabilities</td>
<td>108</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Figure 1</td>
<td>Health system capacity gap</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Supply side dynamics of mental health provision</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Quality Performance Metric Framework</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Daily Captured Demand</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Trend of Captured Demand by day of the week</td>
</tr>
<tr>
<td>Figure 6</td>
<td>Top 4 distributions superimposed on Site A data</td>
</tr>
<tr>
<td>Figure 7</td>
<td>Different types of encounters</td>
</tr>
<tr>
<td>Figure 8</td>
<td>Simulation flow at Site Alpha</td>
</tr>
<tr>
<td>Figure 9</td>
<td>IDEF0 depiction of the Military System</td>
</tr>
<tr>
<td>Figure 10</td>
<td>Adapted IDEF military depiction</td>
</tr>
<tr>
<td>Figure 11</td>
<td>Factors leading to increased demand (Part 1)</td>
</tr>
<tr>
<td>Figure 12</td>
<td>Factors leading to increased demand (Part 2)</td>
</tr>
<tr>
<td>Figure 13</td>
<td>High level view of both stressed systems</td>
</tr>
<tr>
<td>Figure 14</td>
<td>High level view - mismatch loop pushing patients to the VA system</td>
</tr>
<tr>
<td>Figure 15</td>
<td>Downsizing and Loss of Knowledge Effect on Military Health System</td>
</tr>
</tbody>
</table>
CHAPTER I: CAPACITY GAP IN THE MILITARY MENTAL HEALTH SYSTEM

'Capacity has always been one of the most important strategic variables ...'

INTRODUCTION

In 1989, Eppen et al. (p.1) in reference to automobile companies noted that 'capacity has always been one of the most important strategic variables.' Today however, the same could be said of health systems. Dingman, Theobald and Jefferson suggest that quantifying and prioritizing initiatives that impact system efficiency is very difficult without a formal strategy for managing capacity (GE Healthcare, 2012). Similarly, Hall (2013) reports that effective management of capacity by improving the flow of patients to reduce delays, ensuring synchronization of services and demand patterns, and coordinating services with direct patient care has been shown to lead to dramatic improvements in patient outcomes, satisfaction, access to care, and cost of care.

Over the last few years, the U. S. military has been engaged in a War on Terror which has increased the focus on efficiency and cost reduction in the military mental health system. In 2012, President Obama signed an Executive Order to improve access to mental health service for active duty members and for veterans. Two years later, in 2014, the President outlined 19 new executive actions to improve the lives of service members with a focus on improving access to mental health care. Seven of these actions placed a priority on redesigning the mental health system to improve access, while four others emphasized the need to increased awareness of mental health (White House Statements and Releases, 2012; 2014). These actions outline the importance of mental health.

Research suggests that capacity planning is one of the key methods to increase responsiveness to demand and efficiency. This thesis applies capacity planning by (system redesign) to an outpatient mental health unit at a military installation and compares the redesigned system to the existing system. I explain why capacity planning is important to the military and to health systems, and summarize the literature on capacity planning. In addition, I develop a modeling approach to capacity planning for
the Army mental health system and outline performance metrics to be used to compare the two system designs. Although the methods used here can be applied to even larger systems, this research is conducted using data from the outpatient unit of a military installation for simplicity and convenience.

Since the onset of the War on Terror, there have been roughly 2.6 million veterans of Afghanistan, Iraq and other theaters of the global war on terror (CNAS, 2013). Researchers note that deployment increases the risks of mental health disorders among the troops (Hermann, Shiner, & Friedman, 2012; Hoges, Auchterlonie, Milliken, 2006, Kessler et al., 1995). Other researchers estimate that the overall prevalence of mental health disorders is about 31% among troops that had deployed (National Council, n.d.; DoD Task Force, 2007; Hoges et al, 2006; Hermann, Shiner, & Friedman, 2012). There is also demand within the system from those who have never deployed. Of those service members diagnosed with PTSD from 2000 to early 2013, more than 21,784 (~25%) have never been deployed (Fischer, 2013). This high demand for mental health services is important because mental health disorders have been closely linked to suicides, as well as social and occupational impairments (Holmes et al., 2013). In 2012, there were 349 military suicides noted, the highest recorded number of suicides since 2001 (CBS, 2013).

This imbalance in demand and supply in military mental health must be addressed, as it impacts military readiness and overall strategy. Central to any military strategy is the ‘manning’ or the availability of manpower to carryout the plan. Unmet demand for mental health care threatens the readiness of the military as it affects population health. It affects readiness in that it reduces the number of people available to actively serve, reduces the time dedicated to training, and detracts from the available knowledge and skillset base.

Olhager, Rudberg and Wikner (1998) suggest that capacity management is about strategically dealing with dynamic capacity expansion and contraction based on the level of demand. Similarly, Dekkers (2003) reports that decisions on outsourcing, attaining resources, innovation, process development and performance improvement play a role in strategic capacity management by adapting organizational structures. Capacity management has historically been used to balance the differences in supply and demand. In order to effectively apply capacity management techniques to the military mental
healthcare system I must first of all understand how capacity impacts and is impacted by the health care system.

Capacity and the Health System

Figure 1: Health system capacity gap

Figure 1 captures the key dynamics that come into play in health care capacity management. In general, an existing population generates a profile of care needs. Of the population that needs care only a fraction seek it and only a fraction of those that seek it get it (There is also a fraction that do not need care but seek it, I assume this is a negligible fraction). The fraction of the population that receives care is affected by the capacity of the health system. The quality of the system and the treatments administered affect the system outputs and patient outcomes. These in turn affect the population that need care and the population that seek care. In the context of the mental health care system, the dynamics become even more complex. The difference in the care received and the care that is needed is the capacity “gap.” For simplicity I depict a gap however, as discussed below, there are in reality numerous gaps that impact the system.
Population needing care

One of the key challenges in mental health care is correctly identifying and quantifying need for mental health treatment. Several researchers have noted that the epidemiology of mental health is complex (Shapiro et al, 1985; Mechanic D, 2003; Kessler et al., 2005; Silver, Mulvey & Swanson, 2002; Eveland, 1998; Konrad et al, 2009; Faulkner L, 2003). For instance, the incidence of mental health problems is hard to capture due to the subjective nature of mental health disorders, the variety of disorders and their manifestations, the subjective nature of diagnosis and the frequent occurrence of comorbid conditions. In addition, the distribution of mental health disorders is difficult to characterize, because different people have different disorders at different severities (Fox, Merwin, & Blank, 1995; Kerker et al, 2004). The same event might trigger different mental health problems in different individuals and the same treatment could lead to a variety of results, thus making mental health disorders difficult to control. An increase in the awareness of mental health issues and in the number of people seeking care translates into a bigger care seeking population.

Population Seeking Care

Alonso et al. (2004) as part of the ESEMeD project on mental health noted that most people with psychiatric issues do not get the help they need. Research suggests that 50-60% of people who would benefit from mental health care do not seek it (Cooper, Corrigan & Watson, 2003). Awareness, care seeking behavior and the outputs of the health system are three main factors that impact the population that seeks care. The absence of information about indicators of mental health issues such as depression might cause mental health needs to go unidentified. In addition, lack of awareness about the available resources prevents many people from seeking care. With respect to care seeking behavior, Cooper and colleagues (2003) found that people who blamed themselves for their mental health problems were less likely to seek help. Other prominent barriers to seeking mental health care (in military populations) include stigmatization and perception of the system (Hoge et al, 2004). Many service members were found to worry about the impact receiving mental health care would have on their job and chances for promotion.
Changes in access to care impact the number of people that actually receive care from the system

**Health System: The population receiving care**

The number of people that have access to treatment they need is only a fraction of those that seek care (Alonso et al., 2004). Access into the system can be impacted by any one of the three main aspects of the system design: the nature of the system, the workforce capacity and care pathways.

The nature of the system tells if the system (or its subsystems) are ambulatory or not. A system that provides only ambulatory care cannot serve a patient in need of inpatient treatment.

Secondly, because the workforce capacity is defined primarily by the number of providers in the system, their skills and availability, it severely affects the capacity of the system. The literature on the mental health workforce suggests that core mental health disciplines include: psychiatry, clinical psychology, clinical social work, and psychiatric nursing (Duffy et al., 2004; Elisha, Levinson, & Grinshpoon, 2004). Some researchers extend the list of core professionals to include (licensed) marriage and family therapists (Heisler & Bagalman, 2014; Robiner, 2006; Sargeant, Adey, Quinn, & Milev, 2010). These different provider types have different skills, thus the types of providers that exist in the system affect the capacity of the system. Other aspects that impact the overall workforce capacity of the system include vacation or leave of personnel and the time lag experienced when hiring or training providers.

The third system design aspect that impacts access into the system is how the patients are routed in the system to receive care (i.e. the care pathways). This might be shaped by the norms, protocols or standards in the system, as well as the resources in the system (e.g., the number of support staff and how they are used, the equipment and technology used, how triage is conducted).

**System outputs and patient outcomes**

The design of the system affects the quality of the system. The quality of the system and the quality of the treatment administered in the system impact the patient
outcomes and the system outputs. Thus the system design and the quality of care provided are linked. The key variables in the system include the productivity of the system and the responsiveness of the system to changes in the environment. It is however, important to note that the system outputs are difficult to isolate because they are impacted by systems aspects (such as the patient’s experience of care), individual patient aspects (for instance patient compliance) and occupational aspects which may not be directly affected by the system of care.

As mentioned above, the system outputs and patient outcomes affect the number of people that seek service as well as the number of people that need service. Patients that are successfully treated leave the system and create space for other patients to enter the system. Successful treatment of patients increases the productivity of the system and also improves the perception of the system by the community. This could result in more people seeking care. However, if access into the system is not available, the number of people that seek care drops as people become discouraged by the wait. There are several negative impacts of lack of access to care such as increased severity and increased burden, which are notable but will not be discussed here.

**Capacity and the Health System: Supplier Perspective**

![Diagram](image)

**Figure 2: Supply side dynamics of mental health provision**

6
Aside from the patient population, there are also supply side impediments to access. I find that although there may be a large base population initially interested in providing care, schooling concerns such as fees, duration of the training, advertising and selection processes reduce the available number of qualified providers (Roberts & Bandstra, 2012). Of the providers that go through the school system, licensing requirements and aging of the workforce and attractiveness of other professions detract from the number of practicing providers (De Titter et al., 1991; Scheffler & Kirby, 2003; Duffy et al., 2004; Roberts et al., 2013; Kubiak et al., 2012). The access of these providers to patients and vice versa is impeded by geographic distribution (as most providers prefer to live in urban areas), skillset to demand mismatch (for instance, hiring a psychologist where a psychiatrist might be a better fit), hiring processes (such as those in the military that take very long), burn out (due to overwork or other stresses on the provider), vacation and system design. In terms of system design, the system structure, processes, outputs and context can impede access to patients.

**Capacity management process steps**

Discrepancy in the capacity of a health system (or any institution) and customer demand leads to inefficiencies and subsequently unmet demand. Capacity planning enables the system to meet its customer demand. Green (2004) notes that hospitals are overburdened with inefficiencies and delays, making it imperative for managers to “right-size” their systems and manage their resources effectively. She concludes that capacity management is crucial in order to efficiently manage resources and in order to enable patients to receive the right care at the right time. In light of the challenges that exist in impacting the demand for mental health care, it is necessary to focus on impacting the system design as a means of balancing supply and demand. This focus on decisions that impact the overall capacity of the organization is also known as capacity planning. A capacity planning process can be summarized into the following steps.

1. Assess and quantify the future capacity requirements. This includes estimating the required workforce and the projected demand.
2. Evaluate existing capacity and available resources, including funding.
3. Identify the capacity gap and alternatives for closing the gap.
4. Develop performance metrics and compare the alternatives developed. Consider trade-offs for each alternative.

5. Evaluate the performance of each of the key alternatives under different policies and future possibilities.

6. Choose the best alternative making sure that it is feasible, flexible and sustainable.

7. Implement selected alternative. After implementing, periodically reiterate the whole process.

Although the steps proposed seem straightforward, the capacity planning process is rife with challenges.

**Challenges to mental health capacity planning**

One of the prominent themes in the literature on mental health capacity planning is the obstacles encountered (see Table in Appendix A) when planning for capacity. One main obstacle is the epidemiology of mental health problems. This includes problems with identification, prevalence, comorbidities etc. A second obstacle is the distribution of the workforce. Burke et al. (2013) estimated that in 2010, 90% of the existing mental health centers in the U.S. could not fully meet the needs of their patients. Similarly, SAMHSA (2007b) noted that although more than 85% of the federally recognized areas of mental health shortage are in rural areas, most clinically trained providers are heavily concentrated in urban centers. A third challenge is the variation in effort and practice where different providers provide different quality of care, have different skills, case loads and time allocations, and not every one with a license to practice actually provides care. These make it difficult to estimate the number of active providers and their availability.

Aside from the variations across providers, mental health provision is changing and becoming more complicated. That is, different provider types are assuming new roles as there are shortages and surpluses of different types of providers, and training times are variable. An essential part of capacity planning is effectively estimating future capacity requirements. To do this, understanding utilization patterns is important. However, different communities and socioeconomic groups exhibit different utilization and access patterns. Another key obstacle to capacity planning is the set of systems issues such as
funding and poor commitment to data collection. In addition, lack of standardization of the type of data collected and when, of the way mental health should be coordinated, of key terms in mental health capacity planning and of the guidelines for estimating workforce staffing pose problems. Lastly, demographic, social and political differences make it difficult to compare capacity plans drawn for similar populations.

**Workforce Capacity Planning Models**

Despite the impediments to capacity planning, several methods have been used to estimate the workforce required to close the capacity gap. These methods can be grouped into four types of approaches (Fakhri, Sedian and Daviaud, 2014; Sargeant et al. 2010; Konrad et al, 2009). These are need based approaches for workforce planning, demand based workforce planning, benchmarking models and hybrid models for workforce planning.

Need based approaches make use of data on incidence, prevalence and severity of disease in the target population to come up with estimates of need for services. Adapted Delphi methods (use of expert opinion, research based on policy or clinical guidelines) or other estimation methods are then used to approximate the number of providers needed in any specialty to treat the diseases of interest in the target population. Faulkner and Goldman (1997) used a five-step needs based approach to workforce estimation. In the first step, the types (based on diagnosis, impairment and symptom severity) and numbers of patients to be treated were determined. Next, the treatment needs of these classified patients were estimated. The third step involved role delineation for the different providers that were in accordance with the practice and policy guidelines. In step four, the amount of time providers were expected to spend on different tasks based on specific treatment roles was established. The final step was to determine the amount of time to be spent on direct patient care in different settings. Need based estimates often suggest a need for more psychiatrists (Robiner, 2006; Dial, 1998). While this method might be rigorous and thorough, it has shortcomings. First, it requires significant amounts of detail and data that are not always obtainable. In addition, the level of complexity needs to be carefully managed as medical diagnoses are often complex and hard to distinguish and classify clearly. In fact, this is still a work in
progress as the APA continues to refine and classify psychiatric disorders. Furthermore, this method is also hampered by its inability to react quickly to technological advancements. Lastly, this method does not consider costs or available funding. Funding is a key enabler in the execution of most models. Without funding the required workforce often cannot be hired or maintained.

Demand (or supply) based workforce planning approaches are mainly driven by utilization patterns. That is, current trends in demand for services and past data are used to make estimates about future demand. The number of providers required are then approximated based on this projected data. Here, the approximation of the workforce can be based on past data or from expert opinion or other clinical and policy guidelines relevant to the organization. A variation of this approach bases the estimates on the supply of providers. This method is also limited in that it potentially reinforces demand trends that might not be favorable. Secondly, it is based on historical data so it ignores the relationship between supply and unmet demand, as it assumes that the current supply–demand relationship that exists is adequate for mental health. Third, it is not very responsive to changes in future demand or supply. Similar to the need based method, often cost and technology are not accounted for in these models. This method might also require significant amounts of data on utilization.

The third method, benchmarking, sometimes referred to as service target based approach is the simplest and is often used for macro-level analysis. Benchmarking involves estimating the number of providers needed based on a standard in a specified region or system (Moore & Nelson, 1998). Benchmarking permits comparisons with other regions or systems, with the aim of minimizing the workforce without incurring adverse consequences for the population of interest (Sargeant et al., 2010). One of the most common benchmarking methods used in workforce models is provider to population ratios. When benchmarking estimates are based solely on current utilization data and not adjusted for potential future changes, it faces similar limitations as the demand based method. Also, it is often argued (Fakhri, Seyedin & Daviaud, 2014; Eveland et al., 1998; Sargeant et al., 2010) that the demographic, environmental and sociotechnical differences across regions and systems detract from the validity of such a method.
As most of these methods have their shortcomings, many hybrid approaches have been developed to complement the three key methods described above. For instance, the State might adopt a demand based approach that is constrained by its funding appropriations. These constraints might force it to consider benchmarking in addition to the demand based approach. Several authors advocate a method that incorporates the best aspects of two or more methods for a standard workforce planning projection methodology (Moore & Nelson, 1998; Fakhri et al, 2014; Sargeant et al., 2010).

As you might have noticed, most of the research to date focuses on the first three steps of the capacity planning process and then jumps to step 7, implementation. This is problematic as plans when implemented often prove to be inadequate and not sustainable. In order to account for the interdependencies that exist in the system and adopt a holistic perspective of the systems capacity, a modeling approach was chosen. A modeling approach would enable us to more effectively capture steps 4 through 7 of the modeling process. Models are important because they allow us to replicate the system and test out new ideas without actual real life impacts on the system (e.g. Scenario analysis). Thus different decisions can be tried out and their effects on the system estimated before actual implementation. By use of a model one can collect data a lot faster than in the real system and one can collect data on systems that do not yet exist for analysis. Moreover models facilitate discussion between key stakeholders, assisting them to create shared mental models.

Development of a model enables the decision makers to incorporate the best aspects of the three prominent methods used in capacity planning. A model can be built using historical demand data and existing resources to understand the current and past states of the system. The model can then be used to estimate future demand using data on incidence, prevalence and severity of disease in the target population. The model can be used to create and test alternative system designs. A model would provide an easy way to understand dynamics interdependencies in the system and to test out different policies on the improved system design before implementation. Despite all the advantages of models, I note that models can be just as time consuming or even more time consuming than other approaches to capacity planning, and might require large amounts of data. As with any
type of analysis, the intuition of the analyst is very important here. I next suggest several principles to help overcome the obstacles encountered in the mental health capacity planning process. Later I describe a modeling approach that facilitates the development and comparison of alternative system designs for the military mental health system.

CONCLUSION

Capacity management plays an important role in strategy. As a result of the difficulties in controlling the need for mental health care (demand management), I turn to system design as a means of balancing the supply to the demand (capacity planning). In this chapter I have introduced the research on mental health capacity management, discussed the dynamics of mental health care capacity, the challenges in mental health workforce capacity planning, the existing workforce capacity planning methods and suggested principal steps to guide capacity planning. I suggested a model based scenario analysis approach to capacity planning that allows us to evaluate alternative ways of reducing the capacity gap. I also showed how the principles developed to guide planning are applicable in both civilian and military contexts.

Although the goal of most capacity planning processes is to change the design of the system (or some aspect of it) to improve the overall capacity of the system, it is important to evaluate and test capacity plans before implementation. The following chapters will discuss development of performance metrics to guide comparison of system designs and actual model development, to facilitate capacity planning at an intensive outpatient unit at an army installation.
CHAPTER II: PERFORMANCE MEASUREMENT, QUALITY AND CAPACITY PLANNING

‘If you want to manage performance and capacity, you have to measure it…’

INTRODUCTION

Gunther (2007, p. 6) noted that performance measurement is an integral part of capacity planning. Similarly the ESRI Capacity Planning and Performance Benchmark Reference Guide noted that without performance benchmarks it is almost impossible to do effective capacity planning (2009, p.3). Moreover, to compare two systems I have to develop a common metric by which to evaluate them. Here I develop a quality performance metric framework for mental health care and pinpoint the main metrics that would be emphasized in this research on workforce capacity planning.

In the last decade there has been a dramatic growth of interest in performance measurement and improvement of health care quality particularly because of the concerns about the growing costs of care, increased mortality in mental health patients, aging populations and medical and technological advancements (Pincus et al. 2011; Adair et al., 2003), and the awareness of the impact of unintended consequences. Health care quality measures provide an important means of refining and increasing the value of mental health care (Hermann et al., 2006; Porter & Lee, 2013). Several researchers have observed the need for improvement in the quality of mental health care and the need for concrete measures of quality in mental health (Watkins et al, 2011). However, there is evidence to suggest that development of measures in mental health care quality lags that of general care (Masi et al, 2000; Kilbourne et al., 2010). Over the years there has been a long and ongoing debate on how best to measure the quality of care (Faumann, 1990; Ierodiakonou & Vandenbroucke 1993; Brugh & Lindsay, 1996), e.g., should it be based on processes or on outcomes or on structures of care? The fundamental hypothesis at the core of this article is that quality of care (mental health care) is contingent on four key aspects: structure, process, outcome and context of care.
In 1966 Donabedian proposed a structure, process and outcomes based approach to health care quality. Though seminal in its time, Donabedian’s approach to the elements of care was linear as it assumed that structures affect processes, which in turn affect outcomes (Turner, 1989; Brugha and Lindsay, 1996). However, several aspects of his framework have been widely adopted in general health and in mental health care. One stream of research builds on Donabedian’s framework by suggesting that meaningfulness, feasibility and actionability are desirable attributes of measures (Hermann and Palmer, 2002). Hermann and Palmer rightly observe that there is a tradeoff often made between maximizing the quality of measures used and capturing the diverse aspects of the mental health care system in the measure. Nevertheless, their framework is limiting in that it does not simultaneously consider all the different system elements but instead focuses on process elements. Porter, a prolific researcher on health care measures, contends that outcomes are the most important aspect in the health care spectrum as they determine value. He suggests that value encompasses quality and is the patient outcomes per dollar spent. His approach, like Hermann and Palmer’s, is limiting as it overwhelmingly focuses on outcomes and does not include the dynamic relationships that exist among the system elements. In the framework discussed in this paper, I consider that quality and value are tightly coupled and are not mutually exclusive. In fact, the definitions used here suggest that value is a necessary but insufficient condition for quality in health care.

Other researchers (Hermann et al, 2006b; Brugha et al. 1992; Wells and Brook, 1988; Kerr et al., 2003; Wynia, 1996) have used a two-pronged approach by which they study process and outcome measures or process and structure measures or structure and outcome measures. However, few have simultaneously studied all three system elements as part of quality measurement. I propose that the elements of care as proposed by Donabedian should be considered simultaneously, and as dynamic in nature and embedded in a specific context.

The State of Quality Measurement

As alluded to above, the current state of quality measurement is quite siloed and has not caught up to the changes in the health care system. That is, researchers have
focused on disease specific measures (Young et al. 2001; Fullerton et al., 2011, McCorry, 2000; Theis et al, 1995) or on metrics related to one or two of the elements of quality. To date, the majority of the measures that have been used in the literature are process measures (Porter, 2010a; Hermann et al, 2004). In 2010, only five of the seventy-eight Healthcare Effectiveness Data and Information Set (HEDIS) measures were not process measures (Porter, 2010a). Structural measures have largely focused on staff, equipment and facilities but there is a noticeable but gradual shift toward incorporating organizational elements (Glickman et al., 2007). The literature suggests that there is renewed interest in health care quality measurement as indicated by the domestic and international efforts to develop quality measures (Hermann et al. 2006; Horovitz –lennon et al, 2009; Pincus et al, 2007) that add value (Porter, 2010a). Overall, there is strong emphasis on the need to develop or improve health care quality measures as a whole and mental health care measures in particular (Pincus et al., 2011; Adair et al, 2003; Schneider, 1999).

Structure

Structural factors in the health care system include patient characteristics, provider characteristics, the health care institution characteristics, the equipment etc. The literature suggests that very little attention is often paid to structural components of the system when discussing quality. In fact Mitchell (1998) observed that focus has shifted from structures to processes to outcomes over time. Kilbourne et al (2010) interpret the Donabedian structure element to mean the resources and the policies that exist in the system. Similarly, Porter and Lee (2013) hold that the organizational structure of care is an essential part of care delivery. The structure of the system is essential because it sets the stage, i.e., it ensures that the right setting is in place for the care that is needed. Researchers such as those mentioned above consider the structure to be the characteristics of the treatment settings’ services. Until now, this aspect has been considered mostly related to the availability of resources. Here, I argue that the definition of structure most necessarily should also consider the way the resources are used. That is, it is more informative to know the percentage of psychiatrists that are in administrative roles than to know only the number of psychiatrists in the system. Similarly, what fraction of the
normal providers' time is spent on administrative versus treatment purposes? All these impact the access to care within the system and are structural issues.

In the military, I find that there are different health system structures across different installations and the structure of the system changes as time goes on. An important structural aspect is how providers are ascribed to serve the population. Some army units have Embedded Behavioral Health (EBH) systems which are structured such that a team of providers cater for the behavioral health needs of a specific population. This effectively creates one main entry point into the mental health system. Before EBH, most military mental health systems had several entry points to treatment, thus no one provider was assigned to specific populations.

Another aspect related to structure is the type, mix and number of providers in the system. The military system uses a mix of civilian and military providers. These mental health providers can be counselors, psychiatrists, psychologists etc. These providers may also practice collaboratively in teams or solo. Other aspects of the structure that should be addressed include the information infrastructure and provider time allocation (could be impacted by provider role and whether the provider is in the military or not as previously discussed in Chapter 1). Another structural aspect in health systems is the organizational reporting structure. That is, is the organization a matrix structure or functional? In military systems, while civilian providers may be reporting to one line of leadership based on their functions, the military providers have to report to two lines of leadership, one affiliated with their function and the other to their unit.

Processes

Process has been defined as “the content of care, how the patient was moved into, through and out of the health care system and the services that were provided during the care episode.” Other definitions suggest that processes are activities that are performed on the inputs or the patients (Jenkins 1990, Brugha and Lindsay, 1996). Processes also signify the clinical interventions carried out on the patient (Tugwell 1979, Micossi et al, 1993). Process based performance measures are increasingly being used for benchmarking care, for estimating the gap between evidence-based practice and for accountability in health care (Kilbourne et al., 2010). Several measures of care that have
been developed are process driven. This is because it is hard to measure outcomes as they can be affected by any number of factors not directly related to the clinical care provided and also because processes can be directly impacted by providers. Process based measures are thus typical in quality assessment (Lilford et al, 2007). The quality of the process of care is, however, complicated by the absence of standard processes, the diversity in the providers, and the complex classifications of disorders. When developing process measures the strength of the association between the process indicator and the existing structures and the desired outcomes of interest should be considered.

In the military mental health system (as in other health systems) process includes the care pathway, that is, who does the service member have to contact to receive care or to initiate the process? The care pathway in the military usually includes at least one member of the chain of command. Triage is a part of the process as it impacts how long it takes to get patients to where they can receive the right care. The process of care delivery can be assessed based on adherence to clinical guidelines and awareness of military restrictions. Examples of process metrics include: checks that ensure that certain service members are not prescribed medications that impair their ability to carry out military duties; time between appointments (scheduled and walk-in); time to process discharge; time between follow-up appointments; time between prescription and medication receipt; time for information update; the speed of information sharing between providers and departments and the speed of information update between off-post TRICARE providers and on-post providers.

Outcomes

Simply speaking, outcomes are the results of patient interaction with the system of care. There is a marked paucity of research in outcome measures (Rosenbeck & Cichetti, 1998; Masi, Jacobsen & Cooper, 2000). Porter (2010a) contends that outcomes are the most essential aspect of the system of care as this is the primary means by which value can be measured. Outcome measures are generally classified as fatal or nonfatal. The nonfatal cases are hard to measure because the impact of care is not easily separated from external factors that might influence the outcome (see chapter 1). The fatal cases, when they occur, are few and far between thus the data collected is generally not
sufficient for sound statistical analysis. Another approach common in the literature recognizes five groups of outcomes: death, disease, disability, discomfort and dissatisfaction. This is limiting as it does not account for desirable outcomes such as increased functionality and pain relief (Mitchell et al. 1998; Porter 2010a, 2010b). Outcome measures of disorder progression, functioning and interventions are crucial in the system of care. However, challenges to outcome measurement include assumption that the care provided in private institutions is top tier and need not be assessed. Researchers are also deterred by the additional steps needed to ensure that differences in outcomes are not based on clinical severity but on treatments administered. There is also the problem of isolating the system related impacts on the outcome from other environmental and patient related aspects that impact the patient outcome, especially for mental health.

In chapter 1, I divided the outcomes of the mental health system in to system outputs and patient outcomes. Patient outcomes include the patient’s satisfaction with services received, improved health conditions and improved functionality. In the military, separation from service is a possible outcome as well. In terms of system outputs, some indicators include number of people separated for medical reasons, number of fatalities and number of patients treated per time period.

Donabedian (1992) and Brugha and Lindsay (1996), see health status or outcomes as a result of the antecedent health care. This view recognizes a connection between the processes that provide care and the outcomes of the patient. While this approach is sensible it does not inform on how the process and/or outcome measures might be measured. Because of the challenges involved, some researchers focus on outcome measures and consider processes only when adverse events occur. Others look only at the process measures with the assumption that the outcomes would be a reflection of the processes. Jenkins (1990) notes that it is more useful to measure the inputs and the outputs, and processes should only be measured when necessary.

Although it is tempting to focus solely on the outcomes of care, it is necessary and important to include the processes and structures of care that are linked to the desired outcomes. Very few authors have examined all three simultaneously (Hammermeister et al, 1995; Kilbourne, 2010). A rare example where the connection of structure, process
and outcomes is made is in Rogers et al. 1993. Despite these attempts there is still a void to be filled.

Some researchers adopt a disease specific perspective (for instance, number of suicides by Major Depressive Disorder patients) that ignores impacts of the system, while others have been known to focus mainly on patient related measures or health provider measures. These specific measures can be included in the framework provided later in this article to make the measures more explicit and customized.

Along with the fragmented approach to performance measurement, it has also been noted that most researchers and practitioners focus on one level of analysis, usually the individual (patient or provider) level (Edlund et al, 2003; Rose et al, 2011). An extensive review of domestic and international literature on performance measurement by Adair et al (2003) confirms the prevalence of studies focused on one level of analysis. Few researchers touch on individual and system level measures or comparisons with respect to quality (Leslie & Rosenheack, 2000; Druss, Rosenheck & Stolar, 2003). In a diverse setting such as the mental health care system where there is ambiguity and multiple stakeholders, with different ideas of what constitutes good performance (evaluative complexity), it is essential for quality assessments to be based on more than one perspective. For instance, patients' perception of quality may be based on their wait times and their interactions with the staff while providers may judge quality as primarily a function of the outcome. All of these different aspects essentially contribute to the quality of care in the system.

**Systems Approach To Mental Health Care**

Now more than ever, there is a need to adopt a systems approach to mental health care quality measurement. This systems approach to mental health care quality is necessary because mental health is not entirely dependent on a one-to-one interaction with the patient and the provider as might have been the case in the past (Mitchell et al, 1998). In cases where the family of the patient is involved it may even be a many-to-many process. More so, patients often receive different kinds of care from different systems of care at the same time. Porter and Lee (2013) aptly observe that state of the art care practices that call for interaction between providers are replacing discretized
practices. Care delivery depends largely on the ability of the system of care to provide the conditions conducive for the patient’s recovery, as well as the treatment administered to the patient by the providers. A systems approach to quality calls for multiple perspectives and levels of analysis (Watkins et al, 2011; Crawford et al, 2011).

A systems approach to care requires a refinement of the way we think of the elements of quality. Quality has been defined as the extent to which health services increase the chances of desired outcomes and are aligned with professional knowledge (Hammermeister et al. 1995). This definition establishes the link between processes, outcomes and best practices in care but does not clarify the link of structure, process and outcomes to quality. Moreover, the definition assumes that health professionals know which processes and structures lead to improved outcomes. There is limited evidence to support such assumptions. In this section I suggest that quality should also be thought of as having the right structure in place to enable care, and the right processes to ensure that care is provided in the right way such that the right or desired outcomes are obtained for the right context.

Dynamic nature of the elements of the system

There exists a dynamic relationship between the elements of the system (structure, process and outcomes). That is, in effect, the elements of the system are co-dependent.

Quality performance metric framework

The framework presented here builds on the works of several researchers. In 2010, Pincus observed that the framework for the aims of high quality care developed by the Institute of Medicine (IOM) are also applicable to mental health - safe (avoiding unnecessary injury), patient centered (care that is responsive to individual preferences and allows the patient values to guide all clinical decisions), timely (reduce waits and delays for those who give and receive care), efficient (no waste of ideas, equipment or energy), equitable (fair regardless of race, gender, etc.).

As noted above, researchers such as Hermann and Palmer (2002), Arah and his group (2006), Fisher and Rublee (2013) and Porter (2010a) have suggested that measures of performance should be feasible, meaningful (i.e., adds value) and actionable. Porter
suggested that importance to patient (and / or health organization), variability (in outcome dimensions) and frequency are additional criteria to be included when choosing outcome measures. These tenets are captured in what I term practicality.

Regardless of the context, measures have to be practical. The ability to transform the theoretical concepts captured in the measures into account is quite important. This may be done through a combination of indicators. The practicality consideration ensures measures are practicable in the given context or would be practicable in the future based on feasible change. Practicality considerations prompt questions like - is the metric or the to-be architecture meaningful, feasible and actionable? Practicality should be taken into account when developing each measure. Practicality merits significant consideration in any discourse about performance measurement. As practicality is to be considered for every measure and across the different elements and dimensions of quality, it is situated at the vertical column at the left edge of the table in Figure 3.

This research situates the components of performance measures in the different system elements and ensures that they are aligned to the dimensions of quality identified by the Institute of Medicine (Dimensions of Quality column in Figure 2). It takes it a step further, to ensure that the measures are situated in the right context and are practical (side columns in Figure 3). The framework presented here is built using access, productivity and responsiveness as key performance measures that impact the quality of any health care system. Access is linked to quality as failure to receive timely care could exacerbate patient symptoms. In some cases it even leads to a secondary illness. This increased severity of symptoms results in increased resource use and costs. Productivity is considered an indicator of quality in terms of the throughput of the system, the number that leave the system without being seen, and the number of adverse avoidable events such as death to the patient. Responsiveness captures the ability of the system to respond to uncertainties. Responsiveness becomes an important measure of performance when we consider practicality and context. That is, as the environment and context of use changes, how does the system respond? What remains practical as an indicator of quality? Because responsiveness is indicated by the changes in access and productivity (when changes occur to the system or in its environment), responsiveness is represented in a left column
that cuts across access and productivity in Figure 3. Practicality of measures is even more important in the design of capacity of the system as a whole.

<table>
<thead>
<tr>
<th>Performance Measures in research</th>
<th>Structure</th>
<th>Process</th>
<th>Outcome</th>
<th>Dimensions of Quality</th>
</tr>
</thead>
</table>
| 1. Access                        | * Number and type of providers  
* Number of hours of operation of the facility  
* Level of insurance  
* Incidence & prevalence in the population  
* Coverage of the patient | * Time between scheduled appointments  
* Alignment with clinical guidelines  
* Types of services rendered  
* Venue where service was provided | * Timeliness  
* Effectiveness  
* Patient centeredness |
| 1.2 Access: continuity           | * Number of patients who see same provider  
* Number of providers contacted to get an episode of illness resolved | * Number of follow up appointments or calls after discharge  
* Time between follow-up appointments or calls after discharge | * Efficiency  
* Timeliness  
* Patient centeredness |
| 2. Productivity                  | * Left without being seen (LWOBS) | * Throughput of the system  
* Number of patient fatalities | * Efficiency  
* Timeliness |

Figure 3: Quality Performance Metric Framework
Figure 3 shows examples of component metrics that make up access and productivity. Responsiveness can be operationalized as a function of access and productivity.

There are several structural, process and outcome indicators of access to care and productivity. Examples of structural aspects that are practical and aligned with the context of mental health care include: number of providers, types of providers, number of hours of operation of the facility, and level of insurance coverage of the patient. Similarly, time between scheduled appointments can be classified as a process based indicator (Mesquita et al., 2008). Types of services rendered and the venue where services are provided are linked to outcomes (Hermann and Palmer, 2002; Donabedian, 1980,1988; Pincus 2010; Kilbourne, 2010; Jack and Powers, 2009).

Another ‘compound’ indicator of access is continuity (Adair et al, 2005). The concept of continuity can in itself be decomposed to process and outcome level indicators. Examples of these indicators that are practical in the context of mental health include: process indicators such as the number of patients who see same provider and the number of providers contacted to get an episode of illness resolved. Other outcome indicators include the number of follow up appointments or calls after discharge and the time between follow-up appointments or calls after discharge (Hermann and Palmer, 2002; Adair et al, 2005; Aday and Andersen, 1974).

Productivity, the rate at which patients receive treatments and are discharged from the system or subsections of the system, is a key metric used to evaluate the performance of the system as capacity changes. An example of this systematic metric outcome-based indicator of productivity which has been linked to quality (Kilbourne, 2010) is the throughput in the system. Simply put, this is the number of people that go through the system. Numerous studies on the flow of patients have shown that there is a link between flow and the quality of care. Hall (2006) and McLaughlin (2008) note that patients are the key drivers of the system, and therefore impacting patient flow will impact quality, cost and patient satisfaction. It is therefore possible to impact the quality of care by improving the flow through the system (Peck, 2013). Another indicator of productivity is the number of people who leave the system without being seen by the provider (balking from the queue).
Quality Performance Metrics and the Military Health System

Table A: System design and the Quadruple Aim

<table>
<thead>
<tr>
<th></th>
<th><strong>Structure</strong></th>
<th><strong>Process</strong></th>
<th><strong>Outcome</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Readiness</strong></td>
<td>Number of providers</td>
<td>Timelines of care</td>
<td>Effectiveness of</td>
</tr>
<tr>
<td></td>
<td>Number of facilities</td>
<td>Provider skill</td>
<td>care</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Provider stress training</td>
<td></td>
</tr>
<tr>
<td><strong>Population health</strong></td>
<td>Education</td>
<td>Continuity of care by patient provider match</td>
<td>Follow-up treatments and calls</td>
</tr>
<tr>
<td></td>
<td>Preventative measures</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Experience of care</strong></td>
<td>Comfortable setting</td>
<td>Safety measures</td>
<td>Patient perception and use of the system</td>
</tr>
<tr>
<td></td>
<td>Facilitated</td>
<td>Information protection</td>
<td></td>
</tr>
<tr>
<td></td>
<td>communication with patient and care givers</td>
<td>Timeliness</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Convenience of care</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Per capita cost</strong></td>
<td>Waste in the system</td>
<td>Timely care</td>
<td>Efficiency</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Standardized processes</td>
<td></td>
</tr>
</tbody>
</table>

It would be remiss to apply this framework to military mental health care and not talk about the Quadruple Aim. The Quadruple Aim is a key tool that drives planning in the Military Health System. The four constructs of the quadruple aim are Readiness, Population Health, Experience of Care, and Per Capita Cost. Readiness ensures that the military force has manpower that is healthy and ready to deploy, and also has medical support that is ready to deliver care any time, anywhere and for every type of mission. Population Health implies reducing ill health through preventive measures, development of increased resilience and treatment. The Experience of Care construct refers to patient centered care that is at the same time family centered, compassionate, convenient, equitable, safe and of the highest quality. Per Capita Cost establishes that care provision
should be valuable by focusing on quality, eliminating waste, unnecessary variation, and total cost over time.

We find that the IOM dimensions of quality measurement can be mapped to the quadruple aim. Readiness to deploy as a force and to deliver care is strongly related to the effectiveness of care and timelines of care. Effective and timely care produces a medically ready force. However, for effectiveness to exist the providers must be skilled at providing the necessary services. In addition, readiness in the military also requires that providers should be able to function effectively in adversarial, peacetime and unexpected conditions. This attribute of readiness comes through military training.

For a population to be healthy emotionally and physically, care must be equitable and continuous. Equity demands that everyone in the population is treated fairly and equally (based on need) and without regard to race, origin, gender, or socioeconomic status. Continuous care builds resilience through education and preventive measures. Continuous care includes treatment and follow-up after treatment to ensure that the patient attains and maintains good health.

The experience of care for most patients, especially mental health patients plays a key role in the outcome of the care they receive. The experience of care is affected by the timeliness of care, by the safety measures instituted to protect the patients, their caretakers as well as the providers. Safety of the patient includes safety of their information. Patient or family centered care is important for positive experience of care.

The fourth part of the Quadruple Aim is the total cost. Timely care reduces costs as it prevents worsening of patient problems. Efficiency reduces waste and ensures that the resources are optimized. Variance reduction also reduces cost. Financial cost is not explicitly considered in this research, as most of the financial cost is borne by the government and not the individual patients or providers.

It is pertinent to realize that the Quadruple Aim includes and extends beyond the IOM dimensions of quality. It is also clearly evident that readiness, total cost, population health and experience of care are all significantly impacted by the structure of the care system, the process of care within the system, the outcome of the care received and the context of care.
CONCLUSION

In this chapter I provided examples of the key constituents of the framework and suggested some measures that can be readily generalized across disorders and across health care systems. A systems approach to quality assessment provides a level of standardization that allows results to be monitored, tracked and compared. It is pertinent to observe that the metrics discussed in this document are also directly linked to capacity management. For instance, capacity management directly influences the care that is provided and the speed at which the care can be provided. Capacity is intricately linked to the structure, processes and outcomes of health care systems and thus provides fertile ground for future research. However, in light of this research, it is important to note that context affects how quality is defined and operationalized and capacity enables quality care.
CHAPTER III: HEALTH CARE SIMULATION MODELS & MODEL CONCEPTUALIZATION

In chapter 1, I introduced the steps in the capacity planning process. However, I did not dwell on the first three steps as these have been discussed extensively in the literature. In Chapter 2, I developed performance metrics by which one can evaluate the quality of the systems with respect to its capacity. As mentioned before, in this thesis I will explore different scenarios and suggest system designs as a way to impact the capacity of the system. I narrow the system of interest to an intensive outpatient unit at an Army installation (Site Alpha), because of our existing knowledge of the installation, the availability of data from the site and for simplicity. In this chapter I provide a brief literature review on health care models and propose the use of Discrete Event Simulation (DES) for capacity planning at Site Alpha.

Literature Review of Modeling in Health Care Systems Analysis

Using models to conceptualize and reconstruct systems is important in health care capacity management. Health care management involves planning and coordinating several costly and scarce resources, while considering multiple stakeholders with different and sometimes contradictory goals, as well as considering the uncertainty inherent in systems such as these. Despite its complexity, capacity planning in the healthcare system cannot be ignored. Batun and Begen (2013) report that high cost coupled with the forecasted increase in demand make healthcare resource allocation more crucial. Models help provide a means to simplify and visualize the system so as to facilitate analysis.

It is also important to note that despite an upward trend in the number of models that exist, there is a void in the literature of research that focuses on complex integrated or holistic systems (Jun et al., 1999; Fletcher and Worthington, 2007; Vanberkel et al., 2009). Moreover, there is a scarcity of research on mental health systems, especially at the holistic level. This research is a step in filling these gaps.
Queue Models

Starting in the 1900s, queuing theory became more prominent in health care where decisions had been made with little or no quantitative analysis. In 1970, Halpert and colleagues (Halpert et al. 1970) were one of the first groups of researchers to suggest using computer models for resource allocation. In the 1980s, Leff and colleagues (Leff et al. 1985) used linear programing to model health care resource allocation.

The key elements in queuing theory remain the same when queuing theory is applied to the health care system. Queues form when customers (patients) have to wait for a service from a service facility (health care delivery system). Although ubiquitous, queues in health care systems are undesirable. The service stations in a queue, how they are arranged and routing and serving protocols determine how customers (patients) flow through the system. Lakshmi and Iyer (2013) note that queuing models provide insights on the required flexibility of resources, the impact of various service orders or various levels of capacity.

Like Lakshmi and Iyer, several other researchers have conducted reviews on the application of queuing methods to health care. These include: McClain (1976) who examined research on the impacts of bed allocation on utilization, wait time and patient turn-away. Smith-Daniels et al. (1988) presented a limited review of methods used in capacity planning and discussed some ways in which queuing has been applied to capacity related problems. Fomundam and Hermman (2007) summarized the application of queuing theory to health care into three main areas: wait time and utilization, system design and appointment scheduling systems. Preater (2002), Nosek and Wilson (2001) have also published reviews of queuing in health care or pharmaceutical systems. Despite the widespread use of queuing in healthcare systems, there is limited application of queuing theory to mental health. A rare case of this is Koizumi et al. (2005) who found queuing to be a dynamic analytic approach to the Philadelphia mental health system.

Despite the extensive use of queuing theory in health care, it has several shortcomings. Leff (2009) and Gupta (2013) noted that the analytical model had several unrealistic and unacceptable assumptions. Most of the difficulties in his research on capacity allocation were overcome by adding more and more constraints to prevent the optimization model from displaying unrealistic behavior. In the end, “it seemed more
straightforward to employ a simulation model than a severely constrained MHAPLP that, in many ways, was functioning like a simulator” (Leff, 2009, p. 8). Also as analytic models try to capture more and more of the realities of the system, they become analytically complex and discouraging to use because of varying exits and entries from the system and varying queue lengths and multiple stages, partial blocking, and different kinds of wait times to consider. These models fail to capture the joint impact of network service units (i.e. simultaneous processes). In addition, they are not readily translated into discussion with policy makers.

*Choice of DES*

For mental health capacity planning in systems similar to Site Alpha, I suggest employing DES event simulation. The nature of the question being asked as well as the existing data suggested that DES would be the most insightful modeling technique.

Although analytical models are simple and easy to apply, simulation models allow comparison of multiple scenarios and provide both insights and quasi-empirical results. Because they are more adaptable than analytical models, simulations when correctly employed can be used to provide more accurate predictions of system behavior (Utley & Worthington, 2012). Simulation enables repeated analysis of the impacts of different architectures, inputs and policies; it enables use of fewer assumptions than in analytical models; and it is a less costly forum for testing feasible and infeasible alternatives. For instance, the cost of trying out a bad system configuration in simulation is minimal compared to the actual costs in a real life system. In the mental health system, failures can be devastating and can result in loss of life. Although simulation models make it possible to do scenario analysis and to ask what-if questions, simulations are more susceptible to modeling error. It is therefore advisable not to rely solely on simulation outputs but also to compare the outputs with the data from the actual system whenever possible.

In the mental health system, time is a critical property, and therefore a dynamic simulation method should be used. This means static simulations, such as traditional Monte Carlo methods, are inappropriate (Law and Kelton, 2000). Because changes in the health care system are often discrete (e.g., arrival of patients, departure of patients,
movement from one level of care to another), simulation models that incorporate discrete events were preferred. Continuous models are more appropriate for continuous phenomena like water flow or electricity flow. Additionally, because endogenous events such as the service time for each patient or the length of treatment are stochastic, and exogenous events such as patient arrivals or demand for services can be stochastic as well, a stochastic simulation was more desirable for mental health capacity planning, especially since the flexibility and sustainability of the system are of interest.

Even though agent based models (ABM) could be tailored to satisfy most of the criteria listed above, the level of granularity, details and data requirements of ABM do not lend themselves favorably to the mental health capacity problem. More importantly, the challenges in validating ABM made it a less favorable choice. In view of the fact that the data available lends itself more readily to DES, as well as the fact that well established techniques exist for validating DES, DES is a sensible choice for analyzing the mental health system. Hybrid models may also be just as fitting and may impart just as much or more insight than DES models. However, these models are still in their infancy and validation of such models might be problematic. Aside from this, the additional time requirements did not seem favorable.

Simulation Models

Simulation models have gradually infiltrated health care systems analysis. The computer samples values from different probability distributions to plan for patient arrivals and service times and keeps tracks of important statistics. Brailsford and Hilton (2001) suggested that the flexibility, power to handle variety, uncertainty, use of graphical interphases and ease of communication make simulation particularly suitable for health care systems simulation. Banks, Carson and Nelson (1996) found that for complex and realistic models with features such as non-Poisson arrivals and non-exponential services, prioritized routing, and capacity constraints on subsystems, simulation is an attractive option as these models defy mathematical analysis.

There are several streams of healthcare simulation research. One main group of health care research is focused on patient flow. Most of this research, however, is focused on how to reduce patient length of stay. The second stream of literature is focused on the
bed requirements of health care centers. In 1984 and later in 1985, Dumas developed a simulation that used diagnostic groups, gender, admission, time of demand, physician type and types of beds to plan expansion, contraction or reallocation of beds based on changes in demand. Even though there is widespread use of simulation in health care, use of simulation in mental health manpower planning is not as prominent. Several simulation techniques have been widely applied in health care but not all are applicable to mental health care. The following paragraphs discuss the different simulation techniques that are amenable to mental health care delivery.

Table B: Different types of models used in the health care literature

<table>
<thead>
<tr>
<th>System Dynamics (SD)</th>
<th>Agent Based Simulation (ABS)</th>
<th>Discrete Event Simulation (DES)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Description</strong></td>
<td><strong>Models a system as a series of (discrete) stocks and flows (Brailsford &amp; Hilford, 2001; Brailsford, 2007)</strong></td>
<td><strong>Models individuals, interactions between individuals, and interactions with the environment (Maidstone, 2012; Sobolev, 2005)</strong></td>
</tr>
<tr>
<td><strong>Level of Application</strong></td>
<td><strong>Strategic (Dangerfield, 1999; Brailsford &amp; Hilford, 2001; Barnes, Goldman &amp; Price, 2013)</strong></td>
<td><strong>Strategic, tactical &amp; operational (Maidstone, 2012; Macal, 2010)</strong></td>
</tr>
<tr>
<td><strong>Inputs</strong></td>
<td><strong>Rates and flow characteristics, analytic expressions (Barnes, et al. 2013)</strong></td>
<td><strong>Agent characteristics, rules, environmental specifications (Barnes, et al. 2013)</strong></td>
</tr>
<tr>
<td>--------</td>
<td>--------------------------------------------------</td>
<td>-----------------------------------</td>
</tr>
<tr>
<td>Use / Advantages</td>
<td>Provides insight on sources and possible effects of different behavior modes and is a good conversation guide (Dangerfield, 1999) Combines qualitative and quantitative data (Brailsford &amp; Hilford, 2001)</td>
<td>Used to model individual agent response and behavior heterogeneity of the agents Easy to explain (Barnes et al. 2013)</td>
</tr>
<tr>
<td>Shortcomings</td>
<td>Effectiveness in modeling flows (patient) is unknown. (Sobolev, 2005)</td>
<td>Determining the appropriate level of detail is challenging (Barnes et al., 2013)</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Loss of the impacts of stochastic variations and detail. Impact on change is unknown. Not used for optimization or specific predictions. (Brailsford &amp; Hilford, 2001)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table B provides a brief description, level of application, advantages and disadvantages of system dynamics simulations, agent based model simulation and discrete event simulation.

**System Dynamics (SD) Models**

A system dynamics model is made of casual loop diagrams, stocks and flows that are used to explore the interaction between real life systems and high-level policies. It is operationalized by differential and algebraic equations. These models are effective in studying high-level effects, and are especially useful in long-term strategic models in which aggregates can be modeled. They are widely used to understand the sources and possible effects of different behavior modes but their effectiveness in studying policy and organizational change is unknown. It has been used several researchers such as Rogelio Oliva, Jay Forrester, Martin Giese and John Sterman (Oliva, Sterman & Giese, 2003;
Forrester, 1997) to simulate dynamic relationships in order to explore the impacts of different types of interventions, timing, delay and feedback. System dynamics has been used to explore the factors that contribute to delays in accident and emergency units (Lane, Monefeldt & Rosenhead, 2000), for resource provision for sustainable long-term treatment in crisis situations (Lubyansky, 2005), and to conduct a macrolevel analysis of semi-institutionalized and institutionalized mental health systems (Kommer, 2002).

System dynamics enhances understanding of systems and the relationship between its component variables as it leverages both quantitative and qualitative data. It is a great communication tool as it helps clarify mental models and answers questions regarding delays, feedbacks and nonlinearities. It provides unique insights but does not give any right answers. It is often employed as a discussion guide. Since the 1970s system dynamics has been applied to health systems because its ability to capture dynamic complexity makes it attractive for health system analysis (Homer and Hirsch, 2006; Brailsford, 2008).

Nevertheless, there are certain drawbacks to system dynamics. This modeling technique uses mainly aggregate level data and therefore some of the intricacies at the interfaces might be lost. That is, individual properties, histories or dynamics are generally not captured in system dynamics models. Individual properties of the users and the suppliers of mental health care, as well as the properties of the mental health care system play an important role in determining the capacity in any service industry, and especially in mental health care. System dynamics modeling uses global structural dependencies for the aggregate factors. As such, it does not capture variation or uncertainty as readily as some other modeling techniques. Because of the effectiveness of causal loop diagrams in visualizing and clarifying high level dynamics of systems, we use causal loop diagrams to explore different plausible scenarios that could impact the military mental health care system.

*Agent based modeling (ABM)*

ABM is a quantitative method of modeling human behavior (Pearce et al., 2010) at the individual level. In ABM, the behavior of each individual agent in the system is specified and the rules of their interaction are outlined. ABM focuses on the interactions
between individuals, and can also be used to capture the interactions between individuals and the environment.

Agent based modeling (ABM) has recently been gaining momentum in the healthcare arena (Barnes, Golden and Price, 2013). In healthcare delivery, ABM models have been used to model the behavior of health providers as demand levels, patient acuity and other emergency department settings change (Kanagarajah et al., 2006). Ringel et al. (2010) applied ABM to study the impact of policy changes on health insurance markets. Laskowski and Mukhi (2009) discussed using ABM to study hospital staffing levels based on demand. Similarly, Wang (2009) used ABM to identify bottlenecks that affect total patient time spent in the system. Other healthcare applications of ABM include patient tracking (Laskowski et al., 2010) and epidemiology (Carley et al., 2006; Cummings et al., 2004; Barnes et al., 2012).

There are several advantages and disadvantages of ABM. ABM can be employed when multiple different actors interact in many different ways. These models also capture detail at the individual and system levels and are easy to communicate. While ABM offers the advantage of heterogeneity as one can model different types of patients and care providers, there are also several disadvantages. ABM has been used in a variety of industries (performance evaluation in manufacturing, scheduling, information search, etc.). In addition, ABM models easily become complex as the level of detail increases (Barnes, Golden and Price, 2013). Similar to other modeling techniques, it is difficult to generate realistic dynamics for particular scenarios and determining the appropriate level of detail can be challenging. ABM models can be especially challenging to validate especially when quantitative (Maidstone, 2012).

Discrete Event Simulation (DES) Models

DES models are often used to model systems with multiple queues and accurately capture interactions in multiple concurrent processes (Fone et al, 2003). In a DES, time progresses at discrete intervals and events are mutually exclusive (discrete). This approach to time allows for optimal use of data to describe the time to event and its effects in the system. This makes it a good option when entities are subject to competing or numerous risks. Similar to Markov chains, the system is modeled as a group of finite
states with transitions occurring on some events. However, it does not face most of the constraints of the Markov approach (Sobolev, 2005). The key components of DES are: resources, entities that compete for resources, attributes that guide the interaction of the entities and the resources, queues that form as the entities try to access the resources, and time. The outputs of a DES include flow times, time spent waiting in the system (or in part of the system), throughput and resource utilization (costs). Changes in the outcomes listed above have been found to affect patient outcomes via changes in time to treatment and access to care.

DES captures emergent behavior and uncertainty. Therefore, it is attractive for analyzing complex systems such as health care. DES is particularly suited to health care because it is flexible and can be used to depict complex behavior within and between system components and the environment. That is, the concurrent interactions that an individual or a population has with others in the health care system can be simultaneously modeled and analyzed. DES can also account for variation, variety and uncertainty. It can be used when patient characteristics are to be taken into account and if they change over time; it can also be used when history matters and when the decisions made along the way in the simulation are of interest. It can also be used to represent a chain of associated events (Karnon et al., 2012).

Although there has been an expansion in the use of DES, it is still not very common in mental health care systems analysis. Several researchers suggest that DES is not only good for what-if analyses and the design of new health care delivery systems but is also a tool to forecast and assess the effect of changes in patient flow, explore resource allocation and investigate the intricate relationships among different system variables (Jun et al., 1999). Despite the expansion of DES in health care analysis, only a few researchers such as Kim et al. (2013) and Jiang (2010) have applied DES to mental healthcare systems.

Like most simulation systems there are advantages and disadvantages of DES. DES models can capture high levels of complexity and uncertainty, are accessible and easy to communicate. However, DES is heavily reliant on statistics so that parameter estimation might be a problem. It also requires vast amounts of data (Morecroft and Robinson 2005), which may not always be available.
Several comparisons have been made between system dynamics and discrete event simulation (Brailsford & Hilton, 2001; Lane, 2000; Robinson, 2001; Brailsford, Churilov and Liew, 2003). System dynamics is deterministic, does not handle variety and variability very well and validation of such models is contentious. Discrete event simulation on the other hand can handle uncertainty and is dynamic but the data requirements are high and it requires extensive use of statistics. Morecroft and Robinson (2005) find that discrete event simulations can be used for what-if analyses while system dynamics models are more widely used to study the interaction of policies, events and feedbacks. In a more general sense, these authors suggest that system dynamics examines the ‘performance over time of an interconnected system arising from its internal feedback structure.’ DES likewise, explores the performance of an interconnected system over time. However, for DES, the systems of interest are subject to internal and external random variation and variety.

**Hybrid and combined models**

It is becoming increasingly common to combine modeling techniques. A hybrid simulation model is the symbiotic deployment of more than one simulation technique so as to mitigate their shortcomings and enhance their strengths. Hybrid simulations enable decision making at more than one level of analysis depending on the level of detail and the simulation methods used. Borshchev and Filippov (2004) present the differences and similarities between different modeling approaches and suggest that Agent Based Modeling can be used together with other modeling techniques (such as discrete event simulation or system dynamics) as a complement.

Chahal and Eldabi argue that the nature of healthcare calls for integrative modeling approaches, which cannot be provided by SD or DES in isolation but might be accomplished by a combination of both (Chahal and Eldabi, 2008). These authors suggest that hybrid models of DES and SD would be able to capture stochasticity (of and within the system), as well as model continuous and qualitative aspects of the system. Brailsford, Desai and Viana (2010) write about the ‘Holy Grail’ that combines DES and SD. In their paper they present some efforts that have been made in the literature but
conclude that there is still no real study that combines the principles of these two methods really well, no real ‘Holy Grail.’

There are different ways in which simulation models can be combined. The hierarchical format uses two separate models, where the outputs of one become the inputs of the other, creating an input output cycle between the models (Chahal and Eldabi, 2008; Brailsford, Desai, and Viana, 2010). Such models have been applied in manufacturing and strategic architecting, e.g. Venkateswaran et al. (2005). Another type of hybrid modeling is the nested model (also referred to as the process environment). Here two separate models are developed but one is ‘nested’ or enveloped by the other. For instance, DES might be used to model the process of a specific system in a system-of-systems and SD used to model the system-of-systems. The main idea here is that the changes in the system creates ‘chaotic’ behavior in the system-of-systems, which is then adjusted, resulting in changes in the specific system of interest. This cyclic back and forth continues until both the nested and the enveloping model reach some semblance of equilibrium. Brailsford, et al. (2010) provide two examples of such models. Another type of hybrid simulation, the integrated format, aims to integrate two or more modeling techniques such that there is no clear demarcation between the two models and their system elements or outputs. That is, there is no clear distinction between the continuous and discrete parts of the integrated hybrid model.

A combined approach, which is arguably the simplest case of the hybrid simulation, is the dual use of analytic and numerical modeling methods so as to enhance the strengths of the methods and limit the effect of their shortcomings. Combined modeling techniques wed the concepts of both analytic modeling and simulation modeling. In some cases analytic models are developed to validate simulation models and vice versa, while in others, the outputs of one of these modeling methods is an input to the other. Lakshmi and Iyer (2013) find that several researchers have tried to combine queuing theory and simulation models to exploit the advantages of both analytical and numerical modeling methods. They suggest that combined modeling (by use of queuing theory and simulation) is especially useful in health care as it has led to practical insights. Ceglowski et al (2007) use a data mining (clustering) technique to group patients based on the similarity of their treatment and then use DES to simulate bed occupancy. Other

38
researchers that have adopted combined modeling approaches include: Kao and Tung (1981), Tucker et al (1999), Jiang and Giachetti (2008), Cochran et al. (2006a, 2006b), Wang et al., (2008), Karnon et al (2009), van Dijk (2000), van Dijk et al. (2008) etc. See Lakshmi and Iyer (2013) for a more extensive list of authors who have used combined modeling approaches.
CHAPTER IV: DATA ANALYSIS AND SIMULATION OF THE INTENSIVE OUTPATIENT UNIT AT SITE ALPHA, 2009-2010

In the next few paragraphs I describe the data at Site Alpha and the analyses conducted. I discuss how each piece of the data may be related to capacity planning. This section includes the data description, the discrete event simulation model description and outputs and finally the limitations of the dataset.

INTRODUCTION

The data describes ambulatory encounters in an Army facility from fall 2009 to fall 2010. The data set scrutinized had 61,221 observations (patient encounters) over a period of 293 days. An encounter here is taken to mean each individual visit to the health system where treatment was sought and interaction with a provider took place. The days observed were primarily weekdays. The dataset does not include data for some weekends and holidays. This is understandable as it can be assumed that most facilities are closed or experience really low traffic on these days.

Descriptive Statistics of Site Alpha

I parsed the data into the daily, weekly and monthly flow. The average captured demand for each day of the week suggests that Tuesdays see the most traffic. Wednesday is next in terms of traffic. Over the weekend, very few patients show up in the clinic. Weekends were left out in most of the analyses below, because they experience extremely low traffic.
Figure 3 suggests fluctuations in the flow of patients from day to day. This is important because we have to manage the capacity to supply care to match the demand for care. The capacity planner must therefore ensure that days which experience the most patient flow get more providers than other days which see much less flow.

When the demand for each day was mapped for a total of 293 days minus the weekends (i.e. 255 days), the above area chart was obtained (Figure 1). Most of the extremely low traffic days were on Fridays or on Mondays following holidays. From a data perspective this was unusual but reasonable considering the context of the system. Discussions with service members led me to understand that in the military, once in a while the service members must leave for training and sometimes they have half days or holidays as well. As most of these tend to fall on Mondays or Fridays, the data is reasonable. Thus training and holidays affect patient flow in the system.

The data was then visualized in terms of weekday flow across the year, to see if the flow for each weekday was the same through out the year, and if the days were comparable.
As observed in Figure 4, there seem to be some similarities and differences across weekdays in the data. An analysis of variance was conducted to investigate if these differences were random. The one-way ANOVA suggested that the differences between day of the week were not due to chance ($F=102.2$, $p=0.0001$). I then conducted a Tukey’s honest significant difference (HSD) post hoc test to find how the days differed. Tuesday and Wednesday were not statistically different ($p=0.40$) but were different from Monday, Thursday and Friday. Similarly, Monday and Thursday were not statistically different ($p=0.99$). Friday was different from all the other days of the week.

These results suggested that more providers were needed on Tuesday and Wednesday than on other days of the week. Also, in terms of planning vacations and days off, Mondays, Fridays and Thursdays should be used, as opposed to Tuesdays and Wednesdays.

The next step was then to parse the data by month of the year. In the table below, we see the average captured demand per day for each month of the year.

Table C: Average daily captured demand for each month

<table>
<thead>
<tr>
<th>Month</th>
<th>Average captured demand/day</th>
</tr>
</thead>
</table>

Figure 5: Trend of Captured Demand by day of the week
From the table, I noted the wide range in the average daily demand across the months of the year from 198 patients per day to 296 patients per day. This difference suggested a need to test if the daily flow of patients was statistically different from month to month or if this was by chance. I also set out to investigate if there was an interaction between weekdays and months of the year. That is, did the distribution of patient flow across the days of the week change from month to month or was it always as discussed above? From a capacity planning perspective, this is important in the sense that months with the lowest traffic could be used to plan for provider training, while strategies (see chapter 5) could be adopted to increase capacity in months that experience extremely high traffic.

I first conducted a one-way analysis of variance to test for the difference across months of the year. The difference across month of the year was statistically significant at \( p = 0.01 \). A Tukey’s HSD post hoc test was then conducted to identify which months were different. This test showed statistically significant differences between August and April \( (p = 0.016) \), between June and April \( (p = 0.022) \), between May and April \( (p = 0.019) \) and between September and April \( (p = 0.023) \). In effect, it is clear that the flow of patients is statistically different between April and the summer months. One hypothesis to explain...
this difference is the spring break holiday which generally falls at the end of March or the first week of April. If this is the case I expect to see a drastic reduction in the number of patients captured in the system at this time. Another hypothesis could be that the change in the weather to spring temperatures at this point in time affects the number of patients that are captured in the system. If this is the case we expect an overall decrease in the patient flow from March 20 through mid June (spring months in the U.S.A). A close observation of the data suggests that there is indeed a drop in the daily patient flow from the last week of March through April. However, while the data shows no clear dramatic drop that could be attributed to spring break, this observation could not discredit the idea of the impact of weather. A different hypothesis that could apply here is the impact of post changes (PCS) on the utilization, as most post changes are made over the summer months. This increases utilization over this period as service members have to go through a medical check-out and in-take each time they leave or arrive a military post. Even though we note that seasonal temperatures and the timing of post changes (PCS) might have an impact on the patient flow, I do not pursue this path further.

I then conducted a two-way ANOVA to investigate the interaction of day of the week and month of the year. The results show that there was no significant interaction (p>0.1). Therefore, it is reasonable to plan for more capacity on Tuesdays and Wednesdays throughout the year. April should be targeted for long vacations and/or trainings.

Distributions

As I had chosen to use DES to conduct the scenario analysis to plan for capacity planning, I needed to parse the data in ways that are useful for building the model. One pertinent input in the model is the patient flow. To determine the distribution of patient arrivals based on the data set, I created a histogram of the total daily arrivals. The shape of the histogram provided information on the general family of distributions. Various distributions were then fitted and observation and statistical analysis are then used to choose the model with the best fit.
Top four distributions on the same diagram

Figure 6: Top 4 distributions superimposed on Site A data

It is important to note that these diagrams suggest why some of the distributions such as Poisson, gamma, and negative binomial, which have been commonly used in the literature to model customer movement or patient flow, were not suitable here. For instance, in Figure 6 above, the Poisson distribution was left out as it did not capture the tails of the distribution and was heavily centered on the mean (This misfit of the poisson distribution is shown in figure 7). By observation of the chart in figure 6, it appears that the generalized extreme distribution, the t location-scale, the logistic distribution, and the Normal distribution could be used to represent the data. In the simulation, I used the Normal distribution to model patient inflow into the system. The Normal distribution was chosen because of its simplicity and match to both the simulation software and the data.
Another data requirement for the discrete event simulation which I had to estimate was the service times. To do this, I decoded the procedural codes so that we could identify the standard expected completion time for each procedure. Because there are over 70 different procedures recorded and because the decoding is very time consuming, I decided not to use all the CPT codes in the estimate. I found that for each clinic, the top 5 most common procedures (extracted using the CPT1 codes) represented more than 60% of the data. The top 5 procedures were then averaged to provide the values in table G.

Table D: Service times per clinic

<table>
<thead>
<tr>
<th>Time</th>
<th>Psychology Clinic (%)</th>
<th>Psychiatry Clinic (%)</th>
<th>SRP Clinic (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>28.2</td>
<td>40.1</td>
<td>91.6</td>
</tr>
<tr>
<td>45</td>
<td>46.9</td>
<td>44.4</td>
<td>8.1</td>
</tr>
<tr>
<td>60</td>
<td>24.9</td>
<td>15.5</td>
<td>0.3</td>
</tr>
<tr>
<td>Average Time (mins)</td>
<td>44.5</td>
<td>41.3</td>
<td>31.32</td>
</tr>
</tbody>
</table>

Because of variation that results from patient, provider or system characteristics (for instance, a patient that takes up more time than expected, arrives late, human error or unexpected procedures) it would be important to incorporate some level of uncertainty. Also, some patients have more than one procedure done on them when they enter the system so this increases the service times for a fraction of the patients, making it important to incorporate some uncertainty in the time spent with the patient. In the Discrete Event Simulation model, these average times were incorporated. However, I also assumed that the times varied approximately 10% of the time to incorporate some level of uncertainty.

**Distribution of Walk-ins**

A walk-in is a patient who walks into the clinic to seek care without any prior appointment. Although a fraction of the patients are scheduled, the majority of the patients who access the system are walk-ins and a tiny fraction are telehealth encounters.

![Frequency of different types of encounters](image)

Figure 8: Different types of encounters
From Figure 7 above we note that the number of walk-in patients was more than half of all appointments received in the system. Scheduled appointments took up about one third of the appointments and the rest were telehealth appointments.

**Number of providers at each facility**

The number of providers available in each facility for each of the days recorded in the data was estimated. Weekends were discarded. The average daily number of providers available in each clinic is shown in the table below. Note that these are rounded values.

Table E: Number of providers by clinic

<table>
<thead>
<tr>
<th></th>
<th>Psychology Clinic</th>
<th>Psychiatry Clinic</th>
<th>SRP Clinic</th>
<th>All data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Daily average</strong></td>
<td>24</td>
<td>12</td>
<td>8</td>
<td>44</td>
</tr>
<tr>
<td><strong>Daily availability</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>for clinical work</td>
<td>20</td>
<td>10</td>
<td>7</td>
<td>37</td>
</tr>
<tr>
<td>(.85 of daily ave.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Staffing per Clinic (%)</strong></td>
<td>54</td>
<td>27</td>
<td>19</td>
<td>100</td>
</tr>
</tbody>
</table>

The table above shows the average daily number of providers per clinic. Information from subject matter experts suggested that at this point in time, the policy was such that roughly 85% of the provider work time was to be spent on clinical work (thus the .85 of daily average on the table). The table shows rounded values of daily availability of 85% of the providers as input to the model.

**Simulink Discrete Event Simulation (DES) Model Description:**

Simulink software was used to create the DES because of its availability and simplicity. The key assumptions here are that patient arrival times are uniform across each hour, even though the overall distribution of arrivals is normal. In addition, external transition probabilities from the different clinics back to the potential patient population are estimated by trial and error. We also assume that patients continue to be serviced even during lunch hour and that providers in general do not see patients after workday hours.
This assumption was based on discussion with a subject matter expert. For a step-by-step description of the model process and other assumptions see Appendix C.

The Model

Simulation Flow Logic for Site A IOP

Figure 9: Simulation flow at Site Alpha

In figure 8 a logical flow of the process of care is mapped. This is based in part from secondary data from past site visits and from subject matter experts. The patient arrives at the centralized scheduling system and is triaged to the clinics, or they may bypass this step and walk-in to the different clinics. Patients may also receive telehealth treatment from any clinic. The patients after being treated may return to the same clinic or leave the clinic. Upon leaving the clinic they may be routed to other clinics or sent out of the system (external transition). The transition probabilities used in rerouting are presented in the appendix. Two features are shown on the diagram, which are only incorporated in the alternate design – the count visit block and the Intensive care block. These features did not exist in the system in 2009 to 2010.
Model Verification and Validation

Throughout the implementation of the model in Simulink, validation and verification checks were used. For instance, errors were traced and tracked until resolved, the inputs at each step (or functional block) were visualized to ensure that they were correct and their intended consequences aligned to the observations. Discussions with subject matter experts also ensured that the model was logically sound. Lastly, output data from the model were compared to the exploratory data from site A, to ensure that the model was indeed calibrated to Site A as it was in 2009 and 2010.

Input data and transition fractions were calibrated to reflect those from site A. For instance, from site A, approximately 42% of the patients who visited the psychiatry or the psychology clinics had scheduled appointments. This was captured in the model by ensuring that at least 42% of the patients entering the specialty clinics went through the scheduling system.

Verification

1. Ensured that the codes used were consistent with Simulink Matlab coding and that all errors were fixed.
2. Used Matlab user exchange forum and experienced coders to ensure that coding logic was correct.
3. Used charts and visual displays in numerous areas of the model to ensure that each function performed as expected.
4. Traced the movement of entities throughout the model to ensure comprehensive logic.
5. Able to replicate results
6. Successful validation of model

Aside from conferring with subject matter experts to ensure that the story displayed by the model was aligned to the actual system studied, I ensured that the model outputs were comparable to the actual system outputs. The model was calibrated to site Alpha such
that there were 37 providers at all times and 15,568 unique patients (compared to 15,490) in the system.

Results of the simulation

Table F: Simulation outputs

<table>
<thead>
<tr>
<th>Provider specialty</th>
<th>Psychology Clinic</th>
<th>Psychiatry Clinic</th>
<th>SRP clinic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Service times (minutes)</td>
<td>Departures</td>
</tr>
<tr>
<td>Psychiatrist</td>
<td>2</td>
<td>51</td>
<td>12</td>
</tr>
<tr>
<td>Psychologist</td>
<td>5</td>
<td>45</td>
<td>3,530</td>
</tr>
<tr>
<td>Psychiatric nurses</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>General nurses</td>
<td>1</td>
<td>45</td>
<td>178</td>
</tr>
<tr>
<td>Psychoanalyst</td>
<td>1</td>
<td>46.2</td>
<td>114</td>
</tr>
<tr>
<td>Alcohol Counselor</td>
<td>1</td>
<td>45</td>
<td>200</td>
</tr>
<tr>
<td>Social Worker Case Mgr</td>
<td>2</td>
<td>45</td>
<td>1,278</td>
</tr>
<tr>
<td>Social Worker Psycho</td>
<td>2</td>
<td>45</td>
<td>1,825</td>
</tr>
<tr>
<td>Physician</td>
<td>1</td>
<td>51</td>
<td>4</td>
</tr>
</tbody>
</table>
When the simulation was run for one year, I obtained the results in Table F. These results suggest that some providers, such as the psychiatric nurses and the psychiatrists at the psychology clinic and the psychologists and social worker at the psychiatry clinic, are under used or are being used for other purposes other than providing treatment. For example, the Social Worker Case Manager at the psychiatry clinic sees no patients during the simulation. The table suggests to us that the provider distribution could be improved. As the main theme of the thesis is capacity planning the simulation suggests primarily that the provider distribution across clinics should be changed.

**Sensitivity Analysis**

As noted in the previous chapters, one of the main aims of the modeling exercise was to consider different scenarios or actions that could be taken to improve the capacity of the mental health care system. I therefore tested the sensitivity of the model to the number of providers. This is very important as the provider time and skills are the main source of capacity in the mental health system. To begin, I increased the number of psychiatrists in the psychiatry clinic by 1. I then observed the service times and number of departures in the system to see what effect this had. It is very interesting to note that this drastically reduces the number of people that get sent to TRICARE (from 42 to 4). However, over all it leads to fewer people being served from each clinic, and a slight increase in service times. This is perplexing and could indicate that the system gets even more clogged with this addition. It is possible that many patients get stuck waiting and have to reschedule or stay for overtime. Upon tracing through the simulation, I found that the Psychology Clinic and the Psychiatry Clinics are bottlenecks in the system. This is

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>Assistant</th>
<th>Technician</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4</td>
<td>45</td>
<td>904</td>
<td>1</td>
<td>39</td>
<td>2</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>45</td>
<td>904</td>
<td>1</td>
<td>39</td>
<td>2</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>


counterintuitive and confirms that adding providers to a system with the current structure is not the best course of action.

When the model is modified again to include an extra provider (psychologist) at the psychology clinic, I find that this new addition does not perturb the system. In effect, I find that adding a psychiatrist and a psychologist to the system has similar effects to adding only a psychiatrist to the system. However, reducing the number of providers in the system, severely reduces the number of patients that are seen in the system. I also found while calibrating the model and by trying different fractions that the modeled system was very sensitive to the fraction of people that returned to the beneficiary population. This makes sense because if the treatment is effective and people get treated and leave the system, more patients can go through the system.

Table G: Model outputs with the addition of 1 psychiatrist

<table>
<thead>
<tr>
<th>Provider specialty</th>
<th>Psychology Clinic</th>
<th>Psychiatry</th>
<th>SRP Clinic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of providers</td>
<td>Waiting times (minutes)</td>
<td>Departures</td>
</tr>
<tr>
<td>Psychiatrist</td>
<td>2</td>
<td>43.2</td>
<td>9</td>
</tr>
<tr>
<td>Psychologist</td>
<td>5</td>
<td>44.4</td>
<td>2,213</td>
</tr>
<tr>
<td>Psychiatric nurses</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>General nurses</td>
<td>1</td>
<td>45</td>
<td>116</td>
</tr>
<tr>
<td>Psychoanalyst</td>
<td>1</td>
<td>46.8</td>
<td>66</td>
</tr>
<tr>
<td>Alcohol Counselor</td>
<td>1</td>
<td>43.2</td>
<td>120</td>
</tr>
</tbody>
</table>
While several inferences have been made about the system above, it is important to note that the data had several shortcomings. The data does not include information on the costs, amount of time spent in treatment or in the system as a whole. The data does not include the number of hours of operation of the facility, the number of patients who left the system because of death, relocation etc or the actual throughput of the system. As noted above, the throughput of the system impacts what happens within the system and is an important capacity planning metric in this case.

The data does not inform us about the number of providers trained in evidence-based practice or the provider experience, even though the level of qualification and skills might be inferred from the specialty. The provider schedules are not available and difficult to estimate from this dataset. The dataset also does not indicate when the providers are overwhelmed by their responsibilities even though this might have a direct effect on their effectiveness. I also cannot tell the time to first appointment as the data depicts only the actual captured demand. As such, the latent demand or unmet demand is hard to infer from this data without potentially erroneous assumptions. This data does not provide information on the 3rd next available appointment for the patient. The time to first appointment and third next available appointment have been used as measures of access in the literature (Meyers, 2003). The time spent servicing each patient is not noted. Instead this is inferred from the procedural codes, making it hard to consider the variability in service times and their impacts on the system.
To add to this, one cannot easily estimate the number of Full Time Equivalents (FTEs) used for direct patient care as opposed to administrative work. The data set used here does not permit an accurate estimate of the percentage of service members on psychotropic medication or with conditions that limit functionality. Moreover, this data does not contain information on the incidence and prevalence of the different disorders in the population as a whole. It can be used to estimate what the prevalence and incidence might be, but only for the captured demand. The diagnoses that are made and treated out of the system or those patients who decide not to seek treatment are not captured in the system. Most unfortunately, I cannot make any hard judgments about the outcomes of treatments based on this data alone. The data on remission of patient symptoms is currently not available in this dataset. It is also very important to keep in mind that there are other forms of care that exist and that are not captured in the data. For instance, complementary alternative medicine techniques such as yoga or informal counseling which the patient might be receiving concurrently with their clinical care is not captured in the dataset. The dataset does not include information on the number of people who leave the system. People leave the system for many reasons, such as change of post, leaving the military service, deciding not to continue to receive treatment, and death. The number of people in each of these brackets is difficult to estimate.

On the whole, the dataset can be used to provide some useful insights about performance. However, one must take into account the uncertainty that is inherent in the system and that is also in the data. The assumptions made while using this data set are explained. Lastly, it is important to note that this material can be used to provide information about the performance of the system and the entities within the system, but cannot provide sufficient information about the efficacy of the treatments administered.

CONCLUSION

In this chapter I described the data and the simulation. I then presented the logic of the simulation model, its outputs and then discussed the weakness of the dataset.

In terms of performance of the system (access, productivity and responsiveness), I find that increasing the number of providers in the system is not a very effective means of increasing performance. In addition, it takes about seven years to train a psychiatrist or a
psychologist. Instead, time should be spent restructuring the system in terms of the way patients are routed and in terms of the way providers are distributed. With regards to patient routing, in the simulation I noticed that patients have to go through a centralized scheduling process to schedule appointments. This setting is problematic as it uses up provider resources, it would increase the times spent in the system if patients are not routed correctly and it could also increase the waiting time delay. This creates a point for further research, where the model could be tweaked and further developed to test the impacts of having or not having a centralized triage system. As explained above, some providers who received few or no patients could be relocated in the system and put in positions where other providers could use some help or used to create teams of care for special needs patients (this idea is discussed in more detail in chapter 5). The next chapter discusses even more scenarios and provides suggestions on how to improve the capacity of the system without relying solely on hiring new providers.
CHAPTER V: INSIGHTS, POLICY IMPLICATIONS AND CONCLUSION

INTRODUCTION

In this chapter I recap the main thesis chapters, and highlight the key findings and contributions of this research. I also present a policy analysis and opportunities for future research.

Recapitulation of previous chapters

In chapter one, I presented a case for the need for improved capacity planning, and proposed a focus on methods that employ scenario analysis. I then suggested 7 steps, which are pertinent to the capacity planning process. The first three steps dealt with the assessment of capacity requirements by forecasting, evaluation of existing resources, and development of alternatives to close the capacity gap. The fourth step suggested developing performance metrics and comparing alternatives. The remaining steps focused on evaluation of the key alternatives under different policies, incorporating flexibility and sustainability measures and implementing a chosen capacity plan.

This thesis focused more on steps four and five, namely, performance metrics and evaluation of alternatives. In chapter two, I developed performance metrics that are aligned to the quality of the system and that are well suited for capacity planning in health care systems. This performance metric framework also provides a guide to the kind of data that can be collected to facilitate capacity planning and system modeling. To facilitate robust capacity planning in health systems, the planner should collect data on:

1. The number and type of providers, their availability and schedules
2. The hours of operation of the facility
3. Insurance of the patients
4. The incidence and prevalence of the disorders
5. Time between scheduled appointments
6. Time between walk-in appointments
7. Kinds of treatments available and their durations
8. Number of patients who see the same provider
9. Number of providers contacted to get an episode of illness treated
10. Number of follow-up calls or appointments after discharge
11. Time between follow-up appointments or calls after discharge.
12. Number of patients who leave without being seen
13. Patient service times in the system
14. Provider and patient attrition (and fatalities)
15. Patient throughput
16. Cost information if cost is a constraint

In order to model the system so that I can develop and test different capacity improvement actions on the system, I presented a literature review of models that have been used in health care systems in chapter three. I also presented reasons why discrete event simulation (DES) was the most appropriate model for use and provided a conceptual model based on queuing theory.

In chapter four, I explore, analyze and parse the data for the DES simulation. Several insights were gleaned from the analyses and the simulation. The key takeaways from this section were that the existing distribution of providers in the system was not well matched to the demand for care. Adding providers to a clogged system is not the best approach because the providers get caught up in the inefficiencies of the system and the extra capacity they bring makes little or no impact. Aside from this, the distribution of captured demand across days of the week and months of the year is uneven. The system needs to create other means of increasing capacity. I also noted it was especially important that the system be effective. Thus, ensuring that the providers in the system have the right skills and can effectively treat the patients is of paramount importance. As people are successfully treated and leave the system, this creates the opportunity for more people to use the system. In the next section I present different strategies to increase the capacity of the system.

**Capacity Enhancement Strategies**

There are several strategies that can improve the match between capacity and demand in the military mental health care system.
**Minor System Changes to Improve System Capacity**

a) Improving the distribution of providers in the system. As I parsed the data for use in the discrete even simulation, it was clear that there were providers in some clinics who received very few patients. For instance, the two psychiatrists in the psychology clinic barely see any patient flow. One or both of them could be moved to the psychiatry clinic. Similarly, in some of the clinics, some of the lower echelon providers hardly saw any patients while others were swamped. One way to approach this could be to move these providers to other parts of the clinic where they are needed and/or train some of them (especially the lower level providers) in the skills that seem to be most in demand. Simply put, allocate more providers to parts of the system that are experiencing higher demand. Skillset match and provider complementarity are of course encouraged. Stefos et al. (2012) studied mental health provider substitution. They noted that in cases where necessary, nonphysicians could be used to support physicians with varying degrees of complementarity. While all providers seemed to enhance the productivity of physicians, those with similar roles were the most helpful. This implies that most providers can provide some value when cross-trained. However, supervision of cross-trained providers is encouraged.

b) A second useful course of action would be to take advantage of the current pattern of captured demand. This means that the managers should create provider schedules and approve vacation days bearing in mind that there are certain days and periods of the year when the demand of the system is lower or higher than usual. Based on the current data system, it makes sense to encourage employees to take leave in months like April with low patient flow or to take short holidays on Mondays, Thursdays or Fridays but not Tuesdays and Wednesdays, whenever possible. Aside from this, the leadership can reward good performance by giving providers time off on some of the low traffic periods or days.

c) Aside from planning around the pattern of demand, the leaders could also try to influence it (demand management). They can create strategies to spread the demand out. For instance, scheduling intensive therapy on Thursday or Friday would shift some of the demand to a different day of the week. Hostetter and
Klein (2013) reported a similar strategy used at Monmouth hospital. Implementation of this strategy not only helped them to reduce the flow inefficiencies but also helped them to improve their revenue.

d) Another possible way to improve the system would be to reroute some of the patients. An important aspect that stood out while I was analyzing the data was the fact that a few patients seem to account for an inordinate amount of demand in the system. That is, 49.2% of the encounters in the system involved only 9.1% of the patients (high utilizers). A deeper look into the data suggested that these patients were seen primarily by psychiatrists and psychologists. A regression analysis ($R^2 > 70\%$) suggested that patients with four or more comorbidities were high utilizers, and patients with fewer comorbidities who had high acuity were also usually high utilizers. The regression analysis suggested that a main diagnosis of anxiety or PTSD in the presence of other diagnoses was significantly correlated with the patient being a high utilizer. PTSD or Anxiety alone was not a significant indicator. Using these criteria, it is therefore possible to correctly identify these patients early on, and send them to a subsystem where they are treated by psychologists and psychiatrists, away from the main stream of patients. This separation of patients based on diagnosis and acuity would lead to reduced time between appointments for all the patients in the system, while providing these patients the intense care and continuity they need. It is important to note that assigning case managers to these patients may be helpful but that in itself may not be as effective as rerouting the patients, as most of these patients had already been working with a case manager at this time. Using this kind of demand partitioning, Brenner (Robert Wood Johnson Foundation, 2014) assigned high use patients to a care team. He worked with 36 patients who had averaged 62 visits per month and a total hospital bill of 1.2 billion dollars. After the intervention, the patients averaged 37 visits per month and roughly $500,000 in total bills.

Major Changes to Improve System Capacity (Scenario based approach)

So far I have described small-scale changes in the system, which are based on insights gleaned from the data from a subsection of the Army mental health system. The
next few paragraphs use general knowledge of the military system, its environment and uncertainties, to present several feasible scenarios and to propose actions that could be taken to improve the capacity to provide care on a larger scale. This section also extends the boundary of the system to include the Department of Veteran Affairs (VA) and makes extensive use of the "pig in the python" analogy. Demographers have used the pig in the python phrase to describe the big hump in the population age distribution that gradually moves as the "hump" ages. It is often used to describe the "baby boomer" population (Krugman, 2000). In this document, the pig would be used to describe huge increases in volume, and the python would be the military health system ending with the VA.

Figure 10: IDEF0 depiction of the Military System

It is hard to adequately depict the U.S. Military System in a diagram. Traditionally the IDEF0 has been used as a communication tool and as a way to analyze and provide a functional perspective. The basic IDEF0 model identifies Inputs, Controls, Outputs and Mechanisms (ICOMs). The input is that aspect that gets transformed by the function of the system or the organization, the controls are the conditions necessary in order to perform a function (also the constraints) and the mechanisms are those aspects
that enable the function to take place (See Figure 9 for an IDEF0 depiction of the U.S. Military System). The military population is generally composed of new recruits, reserve members and active duty members. This diverse population is guided by a strategic focus from the Commander in Chief and also constrained by the skills base and financial resources at their disposal. Several civilian systems provide additional services in support of the U.S. Military System. This system also depends heavily on a health system to keep its members healthy and ready to function. At any given point in the system, service members are retiring, getting discharged or signing on for new periods of service.

Given that the system does not exist in isolation, it is important to show that the Military System receives and feeds outputs into society as this has important implications.

As Figure 10 shows, a healthy and ready force connects into active duty which is part of the input into the organization. Similarly, new recruits come from society into the Military Service. The Military System releases retired service members and discharges service members into the society. The reserve force is part of the society when not activated and part of the active force when activated.
There are several sources of uncertainty in the Army and in the Military System as a whole. Because these populations are different and have been involved in different activities, at any given point in time their health needs are different. It is hard to predict the amount of demand placed on the health system by any one of these populations at any given time, thus the system naturally has very variable demand.

The new recruits interact with the health system during their initial medical screening at which point they can be deemed medically unfit for military service. Aside from this, a fraction of them suffer from adjustment issues and then interact more with the medical system. The reserves, when a part of the active force, also access the military health system. As the force trains and gets ready to carry out its mission it brings together the different groups of inputs shown in Figures 9 and 10 above. Some of these service members get injuries from training or from non-training related activities and so have to make use of the health system. As part of their military career, some of these service members get deployed, while others do not. Aside from the physical injuries from deployment, there are deployment and non-deployment related psychological issues that also demand the services of the health system. After deployment and sometimes treatment, some fraction of the military force retire normally, others get medically discharged and the others who are still fit for duty continue in their military careers, training and preparing for the next mission.

**Pig(s) in the Python**

Consider for a moment, the chain of events that is described above, the interaction of service members at different stages of their military career with the health system. What would happen to this system if there are: 1) ebbs and flows in the demand as a result of multiple deployments, 2) changed strategic focus, 3) downsizing, 4) baby boomer exodus, and/or 5) interaction effects in the system? In the next paragraphs I use causal loop diagrams to explore these different scenarios.

1) **Ebbs and Flows, the periodic pig in the python**

As numerous groups of service members periodically deploy and return from deployments they have certain needs, which have to be addressed by the Military Health
System. These include the pre and post deployment medical procedures or screenings that each service member has to go through. Aside from this, deployment has been positively correlated with high utilization of Psychological health services and attrition from the military (Hoge et al., 2006). This implies that the health system experiences increased demand during preparation for deployment and also immediately following a return from deployment. This creates ebbs and flows in the demand experienced by the health system.

Figure 12: Factors leading to increased demand (Part 1)

The increased demand leads to increased burden of care on the physicians in terms of the volume of care required as well as the skillsets demanded. This increased burden of care, administrative work and responsibilities increase provider stress. Increased provider stress in turn decreases provider effectiveness. There is literature to suggest that the stress reduces professional effectiveness as it decreases concentration and attention span, encroaches on decision making (Askenasy and Lewin, 1996; Smith, 1990; Lehner et al, 1997). Shapiro et al (2005) and Pastore et al (1995) related stress to reduced ability to create strong relationships with patients. Aside from this, to deal with the increased demand, providers may resort to any of the following: more referrals out of Military Treatment Facilities into the TRICARE network, shorter doctor–patient visits
and higher medical discharge and retirement rates etc. This has a negative effect on both the providers and the patients. For the patients, it leads to a reinforcing cycle, as the treated service members stay longer in the system for more treatments because of worse clinical outcomes (Murray, 2000) or inefficiencies. (Thomas, Guire & Horvat, 1996). For providers, this combined with the workload contribute to high turnover rates and fewer replacement of providers, increasing the shortages and bottlenecks in the system (Kruse et al., 2009). Therefore, as the system demands more and more out of physicians, the physician workload increases and efficiency decreases.

![Figure 13: Factors leading to increased demand (Part 2)](image)

As the system struggles with the demand and provider attrition, more people will be medically discharged. As more and more veterans leave military service, they are released into the civilian system (Veterans Health System), which also gets clogged and backed up from the unexpected increased demand. Instead of just one clogged system, we end up with two downward spiraling systems.
The stress on both systems also has implications for cost, experience of care, readiness and population health, as it affects the quality, access, and throughput of the system.

The ebbs and flows in deployment are also characterized by more training, more deployments, more recruiting and faster promotions, etc. This could lead to more physical and/or psychological injuries and inefficiencies, especially when people move through the different stages of their career faster than they can assimilate the knowledge requirement. As a result, the health system experiences increased demand and possibly lower quality/effectiveness. Increased demand, as shown in the figures above, triggers a host of negative repercussions in both systems. It is important to note that although the VA cannot directly impact the military health system by sending patients back, they both have to compete for providers from the same resource pool and also have to compete with civilian health systems for providers.

2. Changed Strategic Focus (A different kind of pig in the python)

Consider a situation in which the U.S. changes its strategic focus such that service members get deployed on different types of missions. This could trigger different types of
injuries and produce a strain on the health system, as a new skillset might be required of the providers, which was not anticipated ahead of time.

Figure 15: High level view - mismatch loop pushing patients to the VA system

Changed strategic focus could also lead to increased use of the VA as more service members are medically discharged when the military health system does not have the resources to take care of them.

3. Downsizing (As the python loses weight, the pig is captured by the VA)

Here the number of new recruits and the civilian support for the military decrease considerably. The DoD also lets go of more people on retirement. This introduces a slightly different dynamic in the military health system. Downsizing results in decreased knowledge base, as some of the knowledge workers are released. The decreasing force leads to reduced number of providers which could lead to a strain in the health system.
While the military experiences a decrease in volume, the VA simultaneously gains the population that leaves the military. The VA gets the pig, while the other sections of the python lose weight. This increase in the population of retirees and patients and families that depend on the VA creates a mix of young and older service members.

4. The Baby Boomer Effect

The baby boomer effect comes about when many people of about the same age group enter or exit the system in the same time frame. The exodus of a large number of people as they reach retirement at almost the same time triggers a need for replacement of lost manpower, thus an increase in retirement is accompanied by an increase in new recruits. The health system experiences a decrease in staff as well, as some providers retire. This leads to decreased capacity especially as it takes time to hire and train new providers (See knowledge loss effect in figure 7, and capacity gap loop in figure 4).

Here, the system experiences increased variability in demand as the health needs of new recruits who are mostly in their early twenties or younger are different from the health needs of seasoned veterans that leave the system (similar to mismatch depicted in figure 6). The exit of these experienced service members from the system results in a loss.
of knowledge, and potentially decreased ability to treat (and train) service members and increased strain on the VA.

5. Interaction effects of downsizing, drawdown and baby boomer effect.

It is worth noting that multiple changes may occur at the same time. As these changes create several interaction effects, I hesitate to speculate on their effects. Instead I use this as an opportunity to suggest that further research may continue from where I have left off, to develop quantitative system dynamics models from the causal loop diagrams so as to study the different aspects described above.

Recommendations

After looking at all these potential threats to the system and considering the way the system is currently structured, it is evident that the Military Health System (and the Army in particular) needs to be able to scale up or scale down as determined by demand. It also needs to be flexible in its capabilities so that it can meet different volumes of different types of demand. I have already discussed a need to restructure the internal Army system (in this case expanded to the rest of the military where applicable). This restructuring can be done concurrently with attempts to increase scalability and flexibility of the system. The following paragraphs provide three key suggestions for incorporating flexibility and scalability in the system: shifting capability internally, shifting capability externally and training.

Shifting capability internally means that resources should be shifted to different parts of the Military Health System as needed, in order to help mitigate the strain on that component of the system. This might require cross training of different types of providers so that they can assist in different sectors, when they are needed. This gives the system internal scalability – where its individual components can scale down and scale up (with relatively low cost to the system) through redistribution of assets.

External shifts in capability would be when we consider shifting capabilities from outside the system to within the system. The current Military Health System reaches out into the civilian network to supplement its care system. An alternative to reaching out to the network could be building a close relationship and a program with the VA (or even
across services) such that the providers within the VA and the Military Health System are interchangeable. That is, when there is higher demand in the VA than in the Military Health System or when the Military is experiencing periods of low demand, the providers can be sent to the nearby VA stations to work there. If the situation were reversed, providers should also be able to move from the VA to the Military Health System as well. This would require that the Military Health System and the VA develop strict regulations and incentives to guide when and how a provider can be made to move within the two systems, and it would require training of the providers. This would be advantageous in that the two systems would be able to respond faster to changes in demand and the skill, experience and talents of the providers are not lost to due to decreased demand for services. As the two systems can scale up or down by supporting one another, they both gain external scalability. A similar idea was proposed by Colonel Grimes when he wrote “until a single management or governance structure for both systems exists, created and mandated by law, the extent and success of collaboration efforts between the DoD and VA healthcare systems will remain limited ....” (Grimes, 2008 p.1).

Another important recommendation would be to establish a learning environment such that new skills are constantly being learned, old skills updated and new talent is constantly learning from the more experienced staff. Such training should also provide opportunities for cross learning such that providers can perform roles different from their primary roles whenever the need arises. This would lessen the impact of the baby boomer effect on the system and also make the system more resilient to changes in the type and volume of demand. These three recommendations would make the Military Health System more flexible and scalable.

CONCLUSION

So far, I have presented a capacity planning approach, a performance metrics framework, data analysis and a model of Army Site Alpha and then delved into more general scenarios that affect capacity planning in the military as well as suggested some approaches to enhancing capacity. The focus has not been on determining a fixed number of providers but instead to match the capacity in the system to the demand. I have not dwelled on hiring of providers for several reasons. The main reason is the development
times for key providers such as psychologists and psychiatrists is roughly seven years. In a highly variable system such as this, a lot can happen in seven years. Aside from this, I am of the opinion that adding providers to an inefficient system would not increase the capacity but might increase the inefficiencies, as noted in chapter 4.

There are several opportunities for improvement and subsequently opportunities for further research based on this work. First, the data set was very limiting as has been discussed in several sections of the thesis. Aside from this, the robustness of the model could be improved. Although this is a follow-up comment that is in part dependent on the data, other software and more robust assumptions could be employed to build the DES event simulation and improve its robustness. Aside from this, the thesis covered only a few steps of the capacity planning process and did not provide any specific numbers to guide the capacity plan. This research provided insights on the system based on data and analysis of the system dynamics, and suggested ways of framing capacity planning and quality metrics in health systems. Although these are important, and add value at the theoretical and policy level, the absence of hard numbers reduces its value at the operational level. However, future research could build on this foundation to complete the steps of combining the scenarios and testing them on the model or capacity plan of interest. More research is need in the incorporation of scenario analysis and simulation in capacity planning. There is also value to be had from developing hybrid modeling approaches. In addition, the use of more specific, standardized quality metrics in health care systems and in capacity planning is still nascent. Performance should be considered in terms of structure, process, outcomes and context.
REFERENCES


Winter (pp. 1469-1477), IEEE


Crawford, Mike J, Dan Robotham, Thana Lavanya, Sue Patterson, Tim Weaver, Rosemary Barber, T Wykes, and Diana Rose (2011). “Selecting Outcome Measures in Mental Health: The Views of Service Users.” Journal of Mental Health. http://content.ebscohost.com/ContentServer.asp?T=P&P=AN&K=62959726&S=R&D=a9h&EbscoContent=dGJyMNXb4kSeqLM4yOvsOLCmr0yep7Bsr6i4SbKWXWXS&ContentCustomer=dGJyMPGut0%2B2qbFluePfgeyx44Dt6fIA.


Donabedian, A., (1980). The Definition of Quality and Approaches to Its Assessment, Health Administration Press, Ann Arbor, MI.

Druss, B G, R a Rosenheck, and M Stolar (1999). “Patient Satisfaction and Administrative Measures as Indicators of the Quality of Mental Health Care.”

Druss, BG, Bradford D, Rosenheck RA, Radford MJ, Krumholz HM. Quality of medical care and excess mortality in older patients with mental disorders. Arch Gen Psychiatry 2001; 58: 565–72. 52


Grimes, B. (2008). Merging the Military Health System (MHS) and the Veterans Health Administration (VHA) into a Single Governance Structure


Hoge CW, Autchterlonie JL, Millikan CS. 2006. Mental Health Problems, Use of Mental Health Services, and Attrition From Military Service After Returning From Deployment to Iraq or Afghanistan.


Hoge, C. W., Auchterlonie, J. L., & Milliken, C. S. 2006. Mental health problems, use of mental health services, and attrition from military service after returning from deployment to Iraq or Afghanistan. JAMA: the journal of the American Medical Association, 295(9), 1023-1032.


Hostetter M. and Klein S. (2013). In Focus: Improving Patient Flow—In and Out of Hospitals and Beyond


Hutschemaekers, Giel, Bea Tiemens, and Ad Kaasenbrood. 2005. “Roles of Psychiatrists and Other Professionals in Mental Healthcare: Results of a Formal Group Judgement Method among Mental Health Professionals.” The British Journal of


http://www.springerlink.com/content/q55m67703wkOqu3m/abstract/2012.


Lane D.C. (2000). “You just don't understand me: Modes of failure and success in the discourse between system dynamics and discrete event simulation”. LSE OR Dept Working Paper LSEOR 00-34, London School of Economics and Political Science

Lane DC, Monefeldt C, and Rosenhead JV. (2000). Looking in the wrong place for healthcare improvement: A system dynamics study of an accident and emergency department


doi:10.1377/hlthaff.22.5.8.


Conference CD.


89


Patel V, Thara R (Eds.), Meeting mental health needs of developing countries: NGO innovations in India, Sage, New Delhi, India (2003)


RAND (n.d), RAND.ppt Retrieved from PTSI share drive spring 2013.


Rogers WH, Wells KB, Meredith KB, Sturm R, Burnham A (1993) Outcomes for adult outpatients with depression under prepaid or fee-for-service financing. Arch Gen Psychiatry 50:517-525


Ross, Sherwood. (2009) “Most Vets Suffering PTSD are not Seeking Treatment.”


Sheppard, Mike (2012) Fit all valid parametric probability distributions to data. Matlab Central


<table>
<thead>
<tr>
<th>Main Challenge</th>
<th>Descriptions</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epidemiology of mental health disorder</td>
<td>Several types of mental health disorders, with different severities, manifestations and comorbidities. Mental health disorders are hard to diagnose and differentiate. Variation in prevalence of mental health disorders</td>
<td>Shapiro et al. 1985; Mechanic D. 2003; Kessler et al., 2005; Silver, Mulvey &amp; Swanson, 2002; Eveland, 1998; Konrad et al. 2009; Faulkner L. 2003; Fox, Merwin, &amp; Blank, 1995; Kerker et al. 2004</td>
</tr>
<tr>
<td>Variation of Effort and practice</td>
<td>Variation in clinical effort as providers have different caseloads, time allocation, and specialties. Variation in quality of care. Use of non-evidence based</td>
<td>Robiner, 2006; Pineus et al. 2007; Mitchell et al. 2009; Druss et al. 2001; Stein et al. 2004; Watkins et al. 2001; Simon et al. 2001;</td>
</tr>
<tr>
<td>Mental Health care provision</td>
<td>Methods</td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>Disparity in the number of qualified providers and the number of practicing providers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overlapping services, lack of clear role delineation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shortages and surpluses in different professions in the mental health workforce</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variable training times for each provider discipline implies varying market entry times</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time lag between training and supply of mental health providers implies impacts of changes are not immediately obvious</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Changing market conditions and policies (e.g. MCOs, ACA)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMA, 1996:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scheffler, Ivey, Garrett; 1998:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baldwin et al., 2006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heisler &amp; Bagalman, 2014:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duffy et al. 2004: Robiner. 2006:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heisler &amp; Bagalman. 2014:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reif et al. 2012:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wang et al. 2005</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utilization</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variation in utilization and access across demographic and socioeconomic groups</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Systems Issues</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Systems issues such as funding, limited commitment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saxena et al. 2007; Robiner.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data sources</td>
<td>2006.</td>
<td></td>
</tr>
<tr>
<td>----------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Heterogeneity of data sources</td>
<td>Robiner, 2006;</td>
<td></td>
</tr>
<tr>
<td>Imprecision and contradictory information in the data sources</td>
<td>Mechanic D, 2003;</td>
<td></td>
</tr>
<tr>
<td>Lack of data integration across different platforms</td>
<td>Duffy et al, 2004;</td>
<td></td>
</tr>
<tr>
<td>Lack of standardization</td>
<td>Thomas et al, 2009; Konrad et al, 2009; Merwin et al, 2003; Faulkner, Larry 2003; Kessler et al., 2002; Robiner, 2006; Heisler &amp; Bagalman, 2014; Robiner et al., 2006; Sargeant et al., 2010; Hoagwood &amp; Olin, 2002; Fakhri, Seyedin &amp; Daviaud, 2014; Burke et al, 2013; Thomas et al, 2009; Dial et al, 1998; Elisha, Levinson, Grinshpoon, 2004</td>
<td></td>
</tr>
<tr>
<td>Lack of consensus on type of data collected and when it should be collected</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No rubrics for coordinating mental health providers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No agreement of the make up of the MHW</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No standard guidelines for estimating workforce staffing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>There exists several different definitions for key terms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demographic, social and political</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Different regions and areas are</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table H: Summary of capacity planning methods

<table>
<thead>
<tr>
<th>Key Stake holders</th>
<th>Need based method</th>
<th>Demand based method</th>
<th>Benchmarking method</th>
<th>Model based method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients</td>
<td>Patients</td>
<td>Community</td>
<td>Patients</td>
<td>Patients</td>
</tr>
<tr>
<td>Providers</td>
<td>Providers</td>
<td>Providers</td>
<td>Providers</td>
<td>Providers</td>
</tr>
<tr>
<td>Community</td>
<td></td>
<td></td>
<td></td>
<td>Community</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data</th>
<th>Prevalence</th>
<th>Utilization data</th>
<th>Proportions</th>
<th>Utilization data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incidence</td>
<td>Provider time use</td>
<td>(provider, population)</td>
<td>High data requirements</td>
<td>Prevalence</td>
</tr>
<tr>
<td>Low to high data requirements</td>
<td>High data requirements</td>
<td>Usually low data requirements</td>
<td>Incidence</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>High data requirements</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Level of applicability</th>
<th>Strategic</th>
<th>Tactical</th>
<th>Strategic</th>
<th>Tactical</th>
<th>Strategic</th>
<th>Operational</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shortcomings</td>
<td>Focus on need</td>
<td>Historical data</td>
<td>High level</td>
<td>High data requirements</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>No alternative generation and testing.</td>
<td>No alternative demand patterns</td>
<td>Different settings</td>
<td>not comparable</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>No plans for sustainability or flexibility</td>
<td>No alternative generation and testing.</td>
<td>No alternative generation and testing.</td>
<td>No alternative generation and testing.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>No plans for sustainability or flexibility</td>
<td>No plans for sustainability or flexibility</td>
<td>No plans for sustainability or flexibility</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Description</td>
<td>System Dynamics (SD)</td>
<td>Agent Based Simulation (ABS)</td>
<td>Discrete Event Simulation (DES)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>----------------------</td>
<td>-----------------------------</td>
<td>--------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Models a system as a series of (discrete) stocks and flows</strong> (Brailsford &amp; Hilford, 2001; Brailsford, 2007)</td>
<td>Models individuals, interactions between individuals, and interactions with the environment (Maidstone, 2012; Sobolev, 2005)</td>
<td>Models systems as a group of finite states with transitions occurring on some events and is stochastic (Cochran &amp; Bharti, 2006).</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Strategic</strong> (Dangerfield, 1999; Brailsford &amp; Hilford, 2001; Barnes, Goldman &amp; Price, 2013)</td>
<td>Strategic, tactical &amp; operational (Maidstone, 2012; Macal, 2010)</td>
<td>Strategic (Barnes, et al. 2013) Tactical &amp; operational (Brailsford &amp; Hilford, 2001)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Rates and flow characteristics, analytic expressions</strong> (Barnes, et al. 2013)</td>
<td>Agent characteristics, rules, environmental specifications (Barnes, et al. 2013)</td>
<td>Arrival times, entity types, resource scheduling, process flows, queuing parameters (Barnes, et al. 2013)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use / Advantages</td>
<td>Loss of the impacts of stochastic variations and detail</td>
<td>Effectiveness in modeling flows (patient) is unknown. (Sobolev, 2005)</td>
<td>Heavily reliant on statistics as parameter estimation might be a problem</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------</td>
<td>---------------------------------------------------------------</td>
<td>---------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Provides insight on sources and possible effects of different behavior modes and is a good conversation guide (Dangerfield, 1999)</td>
<td>Impact on change is unknown</td>
<td>Determining the appropriate level of detail is challenging (Barnes et al., 2013)</td>
<td>Requires vast amounts of data (Morecroft and Robinson 2005)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combines qualitative and quantitative data (Brailsford &amp; Hilford, 2001)</td>
<td>Not used for optimization or specific predictions (Brailsford &amp; Hilford, 2001)</td>
<td>Quantitative ABM models are hard to validate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Used to model individual agent response and behavior heterogeneity of the agents Easy to explain (Barnes et al. 2013)</td>
<td></td>
<td>Software is not readily accessible (Maidstone, 2012) and is time consuming (Maidstone, 2012; Barnes et al., 2013)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Can be applied to systems made of network of queues (Fone et al, 2003). Accurately captures interactions in multiple concurrent processes (Sobolev, 2005) Addresses detail complexity Accessible Effectively models variability, uncertainty, complexity (Davies &amp; Davies, 1994)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effectiveness in modeling flows (patient) is unknown. (Sobolev, 2005)</td>
<td>Determining the appropriate level of detail is challenging (Barnes et al., 2013)</td>
<td>Quantitative ABM models are hard to validate</td>
<td>Heavily reliant on statistics as parameter estimation might be a problem</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Determining the appropriate level of detail is challenging (Barnes et al., 2013)</td>
<td></td>
<td>Software is not readily accessible (Maidstone, 2012) and is time consuming (Maidstone, 2012; Barnes et al., 2013)</td>
<td>Requires vast amounts of data (Morecroft and Robinson 2005)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

104
Captured demand

Demand that is seen within the system at any given time. Captured
demand provides historical data that can be used to understand the
service needs of the service members that enter the mental health
care system. This data is also useful for describing transitions in the
system.

Internal transition probability

Probability that the patient would move from one treatment track to
another or one facility to another or transitions that stem from a step-
down or step up in the level of treatment and supervision the patient
requires. In this case the patient transitions to another mode of care
in the mental health system

External transition probability

Probability that the patient would transition out of the system.
Transitions out of the system could be broken down into transitions
as a result of completion of service or transitions where patients drop
out of the system of care although service is incomplete

Intermodal transitions.

Patients leave the mode of care but not the mental health care system

Micro-level analysis

Focuses on care within each clinic

Meso-level

Focuses on care between clinics

Table J: Example CPT codes – used to estimate service times

<table>
<thead>
<tr>
<th>CPT</th>
<th>Description</th>
<th>Work</th>
<th>Practice (own)</th>
<th>Practice (other)</th>
<th>Malpractice</th>
</tr>
</thead>
<tbody>
<tr>
<td>99201</td>
<td>Office/ outpatient visit, new pt, min</td>
<td>0.48</td>
<td>.70</td>
<td>.24</td>
<td>.03</td>
</tr>
<tr>
<td>99211</td>
<td>Office/ outpatient visit, established pt, min</td>
<td>.18</td>
<td>.39</td>
<td>.08</td>
<td>.01</td>
</tr>
<tr>
<td>99281</td>
<td>Emergency dept visit</td>
<td>.45</td>
<td>.13</td>
<td>.13</td>
<td>0.03</td>
</tr>
<tr>
<td>99291</td>
<td>Critical Care, first hour</td>
<td>4.5</td>
<td>2.95</td>
<td>1.56</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Table K: Example MEPRS3 Codes

<table>
<thead>
<tr>
<th>MEPRS3 Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BFA</td>
<td>PSYCHIATRY CLINIC</td>
</tr>
<tr>
<td>BFB</td>
<td>PSYCHOLOGY CLINIC</td>
</tr>
<tr>
<td>BFD</td>
<td>MENTAL HEALTH CLINIC</td>
</tr>
<tr>
<td>BFE</td>
<td>SOCIAL WORK CLINIC</td>
</tr>
<tr>
<td>BFE</td>
<td>OUTPT SOCIAL WORK CLINIC</td>
</tr>
<tr>
<td>BFF</td>
<td>SUBSTANCE ABUSE CLINIC</td>
</tr>
<tr>
<td>BFF</td>
<td>OUTPT SUBSTANCE ABUSE CLINIC</td>
</tr>
<tr>
<td>BFZ</td>
<td>PSYCH CARE NOT ELSEWHERE CLSFD</td>
</tr>
<tr>
<td>BFZ</td>
<td>OUTPT PSYCH CARE NOT ELSEWHERE CLSFD</td>
</tr>
<tr>
<td>BGA</td>
<td>FAMILY PRACTICE MEDICINE CLINIC</td>
</tr>
<tr>
<td>BGX</td>
<td>COST POOLS</td>
</tr>
<tr>
<td>BHF</td>
<td>COMMUNITY HEALTH CLINIC</td>
</tr>
<tr>
<td>BHG</td>
<td>OCCUPATIONAL HEALTH CLINIC</td>
</tr>
<tr>
<td>BHH</td>
<td>TRICARE CLINIC</td>
</tr>
<tr>
<td>BIA</td>
<td>EMERGENCY MEDICAL CLINIC</td>
</tr>
<tr>
<td>BIZ</td>
<td>EMERG MED CARE NOT ELSEWHERE CLSFD</td>
</tr>
</tbody>
</table>
Appendix B

Figure 17: Top four distributions for walk-ins

Figure 18: Telehealth distribution of encounters

Transition probabilities from facility to facility

To get the transition probabilities, the PatientID and the corresponding treatment venues they visited was extracted from the data. In this data set, 15,766 (25.8%) encounters were with the psychiatric clinic, 27,516 (44.9%) were with the psychologist...
clinic and 17,939 (29.3%) were at the SRP social work clinic. Of all the patients in the data set, 611 used both the Psychiatric clinic and the SRP, 2,132 used the SRP and the Psychology clinic and 464 used the psychiatric clinic and the psychology clinic. 1,887 patients accessed all three clinics, generating 27,756 encounters, out of a total of 61,221 encounters by 15,490 patients. With the understanding that some of these encounters represented comorbid conditions, the data was observed for patients with more than five encounters in at least one location and for whom the difference between the number of encounters at each clinic was less than 10. These patients were considered as the fraction of patients receiving treatment for more than one condition. With these assumptions 464 patients (10,598 encounters) were found to have comorbid conditions. This conservative estimate is used to represent the fraction of patients being treated for comorbidities. In the model, this is represented as a separate server to appropriately capture the resource use.

The transition probabilities were then calculated taking into account the fraction of people who were estimated to have comorbidities. There was no need to account for the 1,423 who used all three of the services because they represented patients moving from one clinic to the other and it was necessary to count their motion from clinic to clinic instead of ignoring it. For the fraction of patients referred from the psychiatry clinic to the psychology clinic, the transition probability was calculated as:

Total referred from psychiatry clinic to psychology clinic divided by the Total number of patients in the psychiatry clinic (this total excludes the number of patients with comorbid conditions).

Table L: Transition probabilities

<table>
<thead>
<tr>
<th>FROM</th>
<th>Psychiatry</th>
<th>Psychology</th>
<th>SRP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychiatry</td>
<td>0.248179612</td>
<td>0.162190866</td>
<td>0.121909922</td>
</tr>
<tr>
<td>Psychology</td>
<td>0.380629868</td>
<td>0.267941132</td>
<td>0.098503236</td>
</tr>
<tr>
<td>SRP</td>
<td>0.121909922</td>
<td>0.098503236</td>
<td></td>
</tr>
</tbody>
</table>
Appendix C

Simulink Model, Construction and Assumptions

To begin, the model uses a Time Based Entity Generator to generate patients. The patients are generated based on an inter-generation time that is normally distributed. The next block is the Output Switch1. This block ensures that only entities generated in the 8-hour work window span on any given day are included in the system. The assumption here is that mental health patients do not successfully access the system at hours outside of the work window. This is a valid assumption because the system is currently meant to simulate the patients that were captured in the system of care.

The entities or patients are then routed into the RoutingPtsClinic Subsystem. Here the patients are separated based on the type of appointment they need. That is, the patients who need to schedule appointments get sent to the SchedulingSys block. In the SchedulingSys block the patients queue up for provider service at any one of the clinics, depending on their needs. Those who wait in line for 6 weeks (720 hours), are sent into the TRICARE network. The other patients who do not need to schedule appointments (the assumption is that telehealth and walk-in appointments are unscheduled – this assumption is backed by the fact that the distribution of telehealth appointments is similar to that of the walk-in appointments) are then separated to their respective appointment types. Fractions of patients with each of the different appointment types are routed to different clinics. For instance, 1%, 49% and 50% of the telehealth patients are routed to the Soldier Readiness Processing Clinic (SRP), Psychology Clinic and the Psychiatry Clinic respectively.

The patients that are routed to the SRP go through the AppTypeAttributes block where they get assigned a treatment time attribute, based on their appointment type. This separation is important in that it makes it easy to customize or change the times spent in treatment if it does vary differently by clinic and by appointment type. Currently the values used for SRP follow a discrete distribution where 91% of the time it is 30mins and 8.7% of the time it is 45min and .3% of the time it is 60mins.

At the SRP Treatment block, the patients are allowed into the clinic only if within the 8hour work window. They then line up based on a First in First out queue discipline.
In the next step, patients that enter the clinic in the 8-hour work window (note that patients are arriving from other parts of the system to this clinic as well so the 8-hour window is necessary) but don’t make it into treatment are then routed for overtime treatment if they are of the highest acuity level, all the others are sent to the reschedule sink. For those that come within the work window (99% are walk-ins) and receive service they get sent to two main types of providers (psychology social workers or social worker case manager). The data indicates that in some cases a technician may also be used but these have in the past seen very few patients.

Patients leave from the treatment block to the ExitsFromSRP block. Here, a portion of the patients are sent to the Psychology clinic, another to the Psychiatry clinic and yet another back to the SRP. Some patients are not sent to any other clinic but are passed out of the system back to the beneficiary population.

The same sequence of blocks and movements is repeated for the Psychology and Psychiatry clinics but with three main differences. For these clinics, there are a wider variety of providers that serve the patients. That is, at the service station of these clinics, patients can be routed to psychiatrists, psychologists, nurses, the psychoanalyst, the alcohol counselor, social workers, physician assistant and technicians. Secondly, a fraction of the patients that use these clinics are routed to the scheduling system, to schedule appointments in the system.

The Psychology Clinic

From the RoutingPtsClinic block, patients arriving the IOP system are sent to different clinics. One of the main destinations of patients from the RoutingPtsClinic block is the Psychology Clinic. This subsystem starts with the dailywkwindow block that ensures that patients arrive the system within the workday. The patients are then separated based on their appointment type. At the Attribute blocks, the patients are assigned a service time attribute. The current values are based on a discrete distribution derived from the data. There is no initial differentiation between telehealth and walk-in patients. This is justified because they follow the same distribution and the number of telehealth patients at this time in the system was reasonable small. However, the impact of the nature of the appointments in each clinic is captured in its estimated service time.
We also noted that only about 34% of the patients that go the SRP clinic go to the other clinics, as such we route half of the 34% to the scheduling center and the rest are allowed into the rest of the system, without going through the scheduling system. In theory this would represent the walk-in patients.

After being separated by appointment type the patients are then routed to different types of providers. From the data, I was able to understand what fraction of patients at the psychology clinic see a psychologist, or psychiatrist, or alcohol or drug abuse counselor, or nurse, or technician or physician aid or social worker. In all there were 10 different types of providers each patient might have seen. Walk-in patients who are in line for treatment but do not make the work hour window are routed to overtime. At overtime, those with the highest acuity are served by a provider and the others are sent to a reschedule sink.

**The Psychiatry Clinic**

The Psychiatric Clinic in the simulation is set-up such that it is almost identical to the Psychology Clinic in configuration of blocks. The key difference is the service time attribute, which is also based on a discrete distribution obtained from the data, the number of the different types of providers and the fractions of people that exit the system.

In summary, the key features of the model include:

- 3 Clinics
- 8-hour workdays
- Repeat appointments
- Centralized appointment scheduling
- Timeout to TRICARE network
- Prioritization (only for overtime)
- Movement between clinics
- 8-hour window for patient arrival
- Overtime for those walk-ins with high IOP acuity
- Transition fractions to other clinics and out of the system
In this model, no one from the SRP is sent out to the TRICARE network, especially since SRP serves as a soldier readiness checkpoint. This is a service that needs to be kept within house.

- The model is run for one year with a warm up period of 100 hours. One year in the simulation, excluding weekends = 24*255 (actual # of data points captured in the data) i.e. 6120 hours, but because the system has already been in existence for a while before the period of observation, a warm-up period of 100 hours is included.

**Data, Assumptions and Shortcomings of the Model**

At various places in the thesis, several limitations of the data were mentioned. Here we see that they have direct impact on the amount of detail that can be added to the model and the number of insights that can be obtained from the model as it is constructed. The most relevant of this is the inability to completely capture the time in the system. The time in the system is affected by the time to check-in and time for service as well as the time spent waiting prior to or in between services. Here I use estimates with 10% variation to capture this. Another key shortcoming of the data that really impacted the model is the number of people who leave the system on any given day. This is important as it tells one how much space is freed in the system and also is a very rough proxy of the effectiveness of the system. A single factor sensitivity analysis suggested that the model was very sensitive to the transition probabilities. Aside from this, no shows were no explicitly captured in the data. This led to the assumption that reneging and no-shows are very rare and only occur as a result of military duty. In addition, with no real information on group therapy, the number, duration and types of services provided, group therapy was not modeled here. In this model, we assume that at this point in the system it does not represent a significant fraction of the treatments administered and so it not essential. This is an important shortcoming, especially if this model were to be applied to the current system, because group therapy provides service to numerous people at once and increases utilization while optimizing resource use. At this point in time, I also did not have information on the number of slots that are reserved for walk-in appointments at each
As such, in the simulation, I assume that each slot is equally likely to be used for walk-in, telehealth or scheduled appointment. No extra slots are set aside for walk-ins.

The model as built is based on numerous assumptions and as such also has its shortcomings. Most of the key assumptions have been mentioned before but other assumptions that were made include the assumptions that the patients are receiving only clinical care and that patients arrive uniformly throughout the workday. I also make an assumption of unlimited system space capacity. The biggest shortcoming of the current model is its inability to provide insights on time between appointments for each patient who is a repeat visit to the system. Nevertheless, the model is a useful tool for identifying bottlenecks and the effects of different policies on the system.