Applying the Theory of Constraints to Health Technology Assessment

Johan Groop, Karita Reijonsaari, and Paul Lillrank
Institute of Healthcare Engineering, Management and Architecture (HEMA)
Aalto University
Espoo, Finland
johan.groop(a)tkk.fi, karita.reijonsaari(a)tkk.fi, paul.lillrank(a)tkk.fi

***Abstract*—Applying assessment methods commonly used in healthcare, such as randomized, controlled trials, and financial analysis, often proves challenging when assessing technologies that strive to improve productivity. The traditional way of seeking the procurement and implementation of process-improving technology is based on financial analyses. These are often not only laborious, but also based on weak assumptions about the interrelationship between technology, processes, and the institutional environment. There is a need for a more pragmatic managerial approach. This paper suggests an alternative means, based on the theory of constraints (TOC), arguing that the primary focus of technology assessment should be on the ability of technologies to remove or alleviate organizational constraints, in order to increase throughput rather than reduce costs. Ultimately, the latter will happen as a consequence of the former. The suggested approach is illustrated through a case study of a home care organization, in which improved productivity is sought through the implementation of a mobile solution. Although conventional financial reasoning held that the implementation would save time and therefore be beneficial, the TOC approach showed that the implementation would in fact have an adverse effect under the current mode of operation.**

Keywords—Theory of Constraints, technology assessment, home care, healthcare operations management, mobile technology

I. Introduction

Most healthcare providers are urged to improve their productivity to cope with increasing demand under resource constraints. There is a rising trend of seeking improvements in productivity through the use of technology, ICT (information and communication technology) in particular. This paper does not discuss clinical technologies, since their assessment follows a different logic and their methodologies are well developed. Instead the paper focuses on technologies that seek to improve productivity through process improvements (henceforth process technologies), following an OM (Operations Management) perspective.

The number of different process technologies on the market is accumulating fast. Decision makers would like to base their decisions on hard numbers, such as net present value (NPV), cost-benefit, and payback time. Data allowing such analyses are, however, rarely available. Evaluating the financial measures of health technologies properly is difficult, for the following reasons [1]:

- Technology solutions often impact on processes and may alter a service as a whole. Therefore it is difficult to evaluate the influence in advance, particularly if the operational mechanisms of a service are not fully understood. There may be opportunity costs and system effects, such as an improvement in one part of a process being offset by adverse impacts on other parts.
- The gold standard of intervention assessment is the randomized, controlled trial (RCT), commonly used in clinical research (e.g., [2]). RCT requires detailed methodological design, including a statistically significant sample and a coherent control population. It can only be done for one technology at a time; once the technology has been implemented. RCT is often too time-consuming and expensive for fast-paced process technology evaluation.

There is a need for a more pragmatic, managerial approach to evaluating process-improving technologies. Such an approach would need to focus on the ability of a particular technology to move the organization towards its goal. Every organization has a goal and defining it clearly is of great importance. The generic goal of any for-profit business organization is to “make money now as well as in the future” [3], “without violating the necessary conditions of providing a satisfying work environment for employees and ensuring customer satisfaction” [4], e.g., by offering quality products or services. The two prerequisites should not be confused as separate goals. They are threshold conditions which need to be satisfied at least to some minimum level, above which their impact on performance diminishes. The goal, on the other hand, has no upper limit, and it is something that should constantly be pursued.

In public healthcare, defining the goal is more ambiguous, because of the inherent complexities. Several authors have suggested defining the goal of healthcare in terms of different outcome-related measurements [5][6][7][8][9], while others have chosen to define the goal in terms of tangibles, such as money [10][11], and the volume of services produced [12] or patients treated [13]. There is no obvious correlation between the volume of procedures and actual health outcomes; sometimes less is more [14]. However, the question of which volume and combination of treatments leads to the optimal outcome is beyond the scope of OM. Therefore efficiency studies must be based on the assumption that a certain clinically justified level of service of a given quality is necessary, and the challenge is to produce those at the lowest (or optimal) cost.

Taking on an OM perspective, Wright and King [13] build on the generic goal, suggesting that the goal of public healthcare systems is to "treat more patients, better, sooner, now and in the future". The authors choose the definition of Ronen et al. [12], who propose that the highest-level goal of a not-for-profit organization, such as a publicly financed health organization, is to "maximize quality healthcare services provided to its customers, subject to budgetary constraints". Both definitions assume that the two necessary conditions are satisfied. The condition of satisfying customers presumes that the right level and quality of service is provided to patients who need it. The authors also find that both definitions implicitly incorporate the pursuit of better health outcomes.

In order to measure an organization's performance relative to its goal, the goal needs to be translated into operational language through clear, simple and appropriate performance measures [15]. Thus, the goal should be defined as something easily measurable [3]. Financial measures, such as net-profit or ROI (return-on-investment), are affected by operational performance measures; therefore focusing on cost measures without a proper analysis of operational indicators is misleading. A certain technology can promote the operational performance measures derived from the organizational goal.

Organizations should focus on throughput, defined as the rate at which the organization achieves its goal [3]. In a for-profit organization where the goal is to make money, throughput is defined as "the rate at which a system generates money through sales" [3]. In not-for-profit organizations defining throughput is more complex and varies according to the specific characteristics of the organization. For instance, in elective surgery throughput can be defined as procedures performed (output) minus rework. Arguing that "attainment of the financial goal [i.e., making enough money to cover expenses] provides for the realization of the clinical goal" (i.e., high-quality care), Gupta and Kline [10] define throughput as the income generated from patient care.

Throughput is not the same as output, although the concepts are related. In manufacturing a finished product is considered output, which is turned into throughput once it is sold [3]. As services are produced and consumed simultaneously, the distinction between output and throughput is less obvious, particularly if payment is made prior to production and consumption. The practical definition of throughput depends on how we define the goal, e.g., money earned, patients treated, or service time produced. As a crucial operational indicator, it is important that throughput be defined as something manageable by a producer, serving its purpose, and that it is easily measurable. In public healthcare, throughput is therefore often defined as something tangible, such as the number of patients treated [13], services rendered [12], or money generated through patient care [10].

This paper takes a healthcare operations management approach, focusing on the managerial aspects of health technology assessment (HTA). The framework follows the management philosophy of the theory of constraints (TOC), in that it focuses on the impact of technologies on organizational constraints. A constraint is generally defined as "anything that limits a system from achieving higher performance versus its goal" [3]. By exploiting or breaking a system's constraints, performance can be improved. The suggested approach explains how constraints management can be used as a tool for technology assessment.

Section 2 provides an introduction to the theory of constraints. The objective is to clarify the logic behind TOC, and to provide reasoning about why it would be a suitable tool for HTA. First, the TOC philosophy and its main principles are explained, and different kinds of constraints presented. The more traditional strategy of reducing operational expense is contrasted with the TOC approach to maximizing throughput. Section 3 describes how TOC can be used as a tool for technology assessment. Section 4 illustrates this approach using an example from an ongoing case study of a home care unit, regarding a decision as to whether or not to invest in a certain mobile platform-based technology solution. In Section 5, we discuss the limitations of the study, the suitability of the TOC approach, and how it relates to financial analysis. Section 7 provides a conclusion and presents suggestions for further research.

II. Theory of Constraints

Originally developed by Eliyahu M. Goldratt [15], "TOC views every organization as a chain of interdependent events (or processes) where the performance of each event (or process) is dependent upon the previous event" [5]. TOC is based on the assumption that every organization has at least one (but no more than a few) constraint(s) that keeps it from reaching a higher level of performance. Without a constraint the system's performance would be infinite [16]. Although there is a tendency to think of constraints in physical terms, e.g., a lack of hospital beds, inadequate MRI capacity, or a shortfall in staff, it has been shown that the majority of constraints are not physical but policy constraints, such as operational procedure or management policy [11], which in turn may cause resource constraints. TOC argues that the system can only be improved by strengthening its weakest link, i.e., the system's constraint. While an organization may experience several difficulties, some problem (constraint) has to be the most significant for the organization's ability to reach its goal [4]. Therefore any improvement effort should target the system's constraints. In this respect TOC differs from other management philosophies, such as Total Quality Management (TQM) and Just-In-Time (JIT), which consider any improvement in a process as being an improvement of the system as a whole [17].

Figure 1. A three-step process and its capacity and constraints (adapted from Ronen et al. [12])

Figure 1 illustrates the logic of TOC through a simplified example. It shows a three-stage process and the capacity of each stage, as well as its demand and output. The demand on the system is for 100 units (e.g., the treatment of 100 patients). The first department can handle 100 units per day, the second department 50 units, and the final department 75 units per day. Department 2 constitutes the system constraint because it can only process 50 patients per day. The system throughput, or in this case the system output, can only be increased by improving the performance of the constraint [3]. "An hour lost at the constraint is an hour lost to the whole system" [15]. Departments 1 and 3 have an overcapacity of 50 and 25 units respectively per day. Fully utilizing the capacity of either department will not impact on the output of the system, but will build up patient-in-process (PIP) inventory [18] (PIP is the healthcare equivalent to work-in-process in manufacturing). The PIP will lengthen the time it takes for patients to get through the system (lead time), as well as increase the ability of the system to respond to new or altered medical needs (response time). This has a negative effect on patient satisfaction. It may also affect clinical quality and thereby increase the operating expenses of the system. Rather than aiming for greater efficiencies through full resource utilization of departments 1 and 3, the system should aim to keep the production flow at a steady rate (determined by the constraint), by ensuring that department 2 can process its full capacity (50 patients per day).

TOC distinguishes between constraints and everything else, which is referred to as non-constraints. According to the above-mentioned logic, the ability of the system to produce is not improved by increasing the efficiency of non-constraints.

A. *The Five Focusing Steps of TOC*

Much like the other OM philosophies, TOC includes a process of continuous improvement, incorporating five focusing steps [3][19]: "1) *identify* the constraint(s); 2) *exploit* the constraint(s); 3) *subordinate* everything else to the decision taken in step 2; 4) *elevate* the constraint(s), and 5) if a constraint is *broken*, go back to step 1. Do not let *inertia* be the cause of a constraint." Step 2 refers to making sure the constraint is fully utilized at all times (provided it is a resource constraint), and eliminating policy constraints. Step 3 means that the constraint determines the pace of production. The main priority of the non-constraints is to keep the constraint(s) fully utilized at all times, their own utilization being of much less significance. Step 4 means improving the performance of the constraint(s) (e.g., by increasing its capacity, eliminating unnecessary work etc.). Step 5 expresses the need to repeat the process.

B. *The Need for a Stable System*

A widely held assertion of OM is that a swift even production flow increases performance [20][21]. Other widespread OM philosophies such as TQM, Lean, and Six Sigma seek to balance the production flow by *balancing the capacity of the system*, so that every step produces at the same rate [22]. Efficiency is pursued though waste reduction; the capacity of each step should equal demand (e.g., a capacity and production rate of 50 units per department in the above-mentioned example). An even flow is sought by emphasizing the reduction of both internal and external fluctuations, in terms of poor quality, the processing time of each production step, quality, worker absence, lateness etc.

TOC, on the other hand, recognizes that fluctuations can never be completely removed. This is particularly true in open systems, such as healthcare, which includes considerable customer-induced variability [23]. Because of the inevitable existence of fluctuations, TOC argues that a balanced system ultimately becomes inefficient. Fluctuations at steps that have no protective capacity cause inventory to build up, prolonging lead times, and reducing responsiveness etc. [24]. The fluctuations are caused by the combination of two phenomena: 1) internal fluctuations, i.e., the statistically inevitable fluctuations inherent in each step, and 2) cumulative fluctuation, caused by dependent events, i.e., the existence of a process. The fluctuations of each step accumulate, rather than level out, causing greater fluctuations downstream [15].

TOC strives for an even production flow, but by means of an *unbalanced system*. This refers to the existence of at least one constraint. Contrary to the conventional view of constraints – something negative that should be eliminated – TOC regards constraints as pace setters, around which the rest of the system should be managed, or *subordinated*. The basic premise is that, by managing the whole system according to the production rate of the constraint, a stable flow can be reached (as opposed to by balancing the capacity and production rate).

According to TOC, before the performance of a system can be improved it must first be stable. This requires the constraint to be identified, so that the non-constrained steps can be subordinated to the constraint, creating an unbalanced stable system. If no constraint exists, TOC suggests creating one, preferably one that is as natural as possible, e.g., an expensive piece of equipment. The positioning of the constraint is an important strategic decision [25].

C. Traditional Focus on Cost

Many healthcare organizations tend to follow a cost reduction policy, seeking to reduce all costs while simultaneously improving the efficiency of resources [5]. Although this is an intuitively logical approach, practice has shown that the measures, decisions, and behaviors this approach leads to can have quite the opposite effect [3][12][15][11]. If greater efficiency of all the steps in a process, instead of the constraints, is sought, PIP builds up, leading to longer waiting times, lead times, and response times, which in turn increases operating expenses. Traditional cost accounting creates incentives that encourage high efficiencies [3]. One reason for is that it is based on the false assumption that all resources are fully utilized at all times [15].

Cost accounting was first developed to deal with the accounting complexities stemming from the rise of mass production in the 1920s. At the time "direct wages were a major component in costs and were considered real variable costs […] The indirect costs were low and assigning them to products based on real direct wages (or another volume-based variable such as machine hours) was a reasonable approximation for making business decisions such as continued production of a product, investment decisions, buy-or-make decisions, and so on" [12]. In those days, indirect costs (i.e., overhead) constituted around 5-10 percent of the total costs of production. Today the indirect costs are 20-80 percent. Assigning them to products or services on the basis of the cost accounting formula creates severe distortions [3][12][15]. For instance, it fails to distinguish between constraints and non-constraints, creating incentives to improve the efficiency of non-constraints, which will not increase the throughput of the system as a whole. In fact, it may cause throughput to decrease, as the flow becomes unstable. This is also known as "the efficiencies syndrome" [12]; a situation where increasing the efficiency of every production step, "emphasizing the utilization of inputs instead of focusing on outputs", reduces the efficiency of the whole system.

D. Focus on Throughput

TOC holds that the most important performance measure of an organization is its performance relative to its goal, i.e., the throughput in terms of the number of "units of the goal" [5]. To improve performance, TOC focuses on maximizing throughput rather than reducing operational expense. This is also known as "throughput orientation" [4][26] or "throughput-world thinking" [26]. The logic is based on the assumption that the change in operational thinking will lead to behavior and decision making that reduces costs, through reduced lead time, response time, and PIP, as well as improved service quality [12]. Operating expenses can only be reduced to a certain level, while throughput, in theory, can be increased infinitely [3].

As the production rate of the primary constraint determines the throughput of the system, the cost of the constraint is the cost of the entire system divided by the available number of constraint hours [15]. This approach is also referred to as throughput accounting (TA) [3][27]. It reduces the complexity by focusing on the production capacity and cost of the production step that matters, i.e., the constraint, instead of all steps separately.

III. An Alternative Approach To Health Technology ASSESSMENT

When health technologies are being assessed, the focus should first be on the technologies' impact on throughput, then on costs. When estimating the time-saving effect of a new technology, it is tempting to evaluate the saved cost in terms of cost per hour. If the time was saved at a constraint this approach might be valid. However, if the time saved does not affect the capacity of the constraint, it is unlikely to increase throughput [1].

Most healthcare organizations have large fixed costs, mainly the cost of personnel. Typically, only 10-15 percent of the budget is variable costs [3]. Saving the time of non-constrained labor (mainly a fixed cost) does not affect unit costs, unless the time saved can be allocated elsewhere (e.g., moving excess capacity to the constraint to increase throughput) or less labor is required as a consequence, while still maitaining full utilization of the constraint. Unit costs, on the other hand, can be reduced by improving the throughput of the organization, as this spreads the fixed costs of the organization over more produced units.

By concentrating on assessing the impact health technologies have on the system's constraints, and thereby on throughput, it is possible to avoid some of the complexities associated with technology assessment in healthcare. If throughput can be increased at a reasonable cost, the additional expense is likely to be more than offset by the increase in throughput.

The first step is to identify the system's constraints using the five focusing steps [3]. Second, the effect of the alternative technologies on the constraints is evaluated. Which constraints does a certain technology affect and how does it improve throughput? Will the technology remove or alleviate the constraint to allow increased throughput? Does the implementation of a certain technology, and the accompanying process changes it makes possible, affect the constraint directly, or only in combination with some other technology or improvement effort?

If it is determined that a technology can remove or alleviate a resource constraint (a policy constraint is not likely to be affected by technology), the marginal return on an investment in the constraint should be examined [25]. In other words, how much additional throughput can the investment achieve? Here, again, focusing on the constraint can reduce complexity.

When assessing different technologies it is not uncommon to encounter a situation where there are several constraints, as well as alternative technological solutions that may affect more than one constraint. Such a situation is illustrated in the following case study.

IV. THE STUDY

This section descibes the TOC approach in use through an empirical example. The investigation illustrated in this paper was performed as part of a larger ongoing study, aiming to explore ways of improving productivity in home care.

A. Aim

Prior to the investigation the subject organization was seeking to improve productivity through the implementation of mobile ICT technology. The objective was to 'save time' on a scarce resource, caregiver time, to be used for coping with an increasing demand for services. With the use of a mobile platform-based technology (henceforth 'solution'), certain office tasks (e.g., charting) were to be transferred to the field, speeding up and improving the efficiency of the caregivers' administrative work routines.

In order to avoid investing in a solution that would not increase throughput, while simultaneously increasing operating expenses, the authors sought to investigate the potential solution's effect on system constraints.

B. Home Care Operations

Home care (or domiciliary care) refers to regular healthcare or supportive services provided in a customer's own home by a visiting caregiver. Home health care, meaning skilled nursing, is sometimes distinguished from home care, meaning non-medical care. Here they are treated jointly under the term home care. The services range from medical (e.g., health care and hospice) to supportive (e.g., bathing) and social services (e.g., transportation). The purpose of home care is to enable people in need of assistance to continue living in their own homes by improving, maintaining, or reducing the natural stagnation of their health conditions and autonomy. Although age is not a basis for service discrimination, the vast majority of the clientele is typically made up of older adults, whose autonomy is reduced.

The frequency of home visits varies with customers' specific needs, ranging from occasional monthly visits up to 4 visits per day, averaging 1.56 daily visits on weekdays in the organization that was studied. Although the absolute majority of the customer encounters constitute home care visits, occasionally services are provided over the phone, or a customer visits the office.

The demand for services is based on an evaluation, performed by the caregivers, typically a nurse and a social worker, in collaboration with the potential customers. During the process, a care plan is made out for the accepted customers. The care plan is revised biannually or when changes in a customer's condition so require.

The output is the volume of procedures or service units performed (the provider perspective). An outcome is a change in a patient's medical condition, which carries some health value (the patient perspective). The outcome (or value) is therefore the ultimate goal. The outcome cannot, however, be used as an operational indicator in situations where the outcome is significantly affected by factors beyond the service provider's control, such as patients' health behavior, placebo effects, or random events. The throughput is therefore the provider's contribution to the creation of outcomes (health value). It was decided that the most appropriate definition of throughput was not the number of customers served, but the service time produced. Defining throughput as the number of customers or visits may risk constructing incentives to maximize the volume of encounters, rather than the effectiveness of the service.

The schedule of the caregivers is based on the demand, as stated in the care plan. In home care, the demand for services is typically greatest in the morning [28]. The caregiver capacity is staggered into two 7.5-hour-long shifts. The morning shift workers start around 7-8 a.m., finishing around 3-4 p.m., while the evening shift staff comes on after 2 p.m., finishing as late as 10 p.m. Although there is a slight variation in the starting and finishing times, the caregiver capacity is quite stable throughout the shifts, with a clear drop between the morning and evening shifts, the capacity of the evening shift being approximately 25% of that of the morning shift.

C. Methods

The inquiry adopted both qualitative and quantitative methods. The qualitative part was conducted to gain a rich understanding of the service operations of the home care organization that was studied. It included interviews and regular meetings with management, an ethnographic study of the service process and work routines, and several workshops with staff. Some results from the qualitative part of the study were first presented in Groop et al. [1].

The management team that was interviewed included the head of Home Care, a service manager, two ICT specialists, and one financial specialist. The objective was to get an overview of the service production system and its dynamics. Once the quantitative study was in progress, regular meetings were held with the management team for the purpose of feedback and analysis of findings. For the ethnographic study, the first author observed a staff member at work for an entire shift. A total of three staff members were observed on three separate occasions. The first two were registered nurses working in separate teams, while the third was a foreman. The first workshop was held at an early stage of the investigation. Its objective was to strengthen the comprehension of everyday services through a facilitated discussion with representatives from all levels of the home care organization. Once the quantitative study was finished, four workshops were held to disseminate and validate the findings. The first two workshops included all the foremen, while the latter two targeted two separate home care teams, chosen by the management.

The purpose of the quantitative part of the study was to quantify the home care operations, focusing on throughput and caregivers' workload. The investigation consisted of a longitudinal analysis of operational data. Data illustrating the distribution of throughput, i.e., the amount, duration, and timing of realized home care visits, were collected for a period of six weeks, from February 9th to March 22nd 2009. This enabled analysis of the caregivers' workload, i.e., services performed in terms of time, over a period of time.

The data consisted of two consecutive three-week scheduling periods, and they were chosen in collaboration with management. According to the management team, the chosen period exemplified the most "normal" conditions, as it was least affected by staffing exceptions as a result of holidays. The data represented a total of 43,716 home care encounters amounting to a total of 21,934 hours of service.

D. *Participants*

The organization that was studied, Espoo Home Care (EHC), is a large public health organization. It is responsible for providing statutory home care services for the City of Espoo, one of the largest municipalities in Finland. The organization is divided into service homes and field-based services (Regional Home Care). Since the potential technology implementation targeted only the field-based services [29], the service homes were excluded from the analysis. Some services, such as night-time care (10 p.m. to 7 a.m.), customer transportation, and meal and grocery deliveries, as well as care for certain severely disabled customers, are outsourced. These services were also excluded from the analysis. Services performed by temporarily leased caregivers, covering for absent employees, were included.

Regional Home Care employed a field-based staff of 326 (as of Feb 27th 2009), of whom 53 (16.2%) were part-time employees, and 19 primarily office-based foremen, who were trained registered nurses or social workers. Out of the field-based personnel, roughly 23% were registered nurses, 61% practical nurses, and 15% home care assistants. The home services provided during the six-week period included services rendered to 2587 customers, by 293 caregivers, for an average of 801 customers per day (STD 40) on weekdays, and 389 (STD 12) on weekends. This corresponds to 1189 (STD 55) customer encounters on weekdays and 671 (STD 44) on weekends.

The clientele of ECH consisted of both temporary customers, who exit the system once their health/autonomy improves, and regular customers, exiting the system either through death or transfer to a more comprehensive form of care, e.g., sheltered accommodation or long-term care. During the period of the study temporary customers received only 0.34% of the total services.

V. FINDINGS

This section first explains the process changes that certain technologies make possible, and provides reasons why these changes might help to improve productivity. After this, the TOC framework is adopted to evaluate whether the suggested process changes would in fact have the intended effect on productivity.

A. *Time and Location Constraints*

The home care service delivery process is field-based, which implies that there are constraints related to time and location [30], which have a great impact on the production flow. In the ECH the caregivers visit the office each morning, to collect their work list specifying which customers to visit, when, and which types of services to perform. The caregivers also pick up the keys to the homes of several of their customers, and company cars, if assigned one. Once the caregivers have completed their rounds, they return to the office to bring back the customers' house keys and to perform office tasks, such as data entry into the EMR (electronic medical record) (i.e., charting: visit summaries, updating patient information and care plans, placing customers' grocery and meal orders etc.). All of these activities are subject to time and location constraints that can be broken using technology. The time constraint means that certain tasks have to be performed in a certain order (e.g., getting the keys before entering a customer's home), and at certain times when the customer and caregiver meet. The location constraint means that certain services need to be performed at locations where the customers (home) or equipment (office computers) are located. The time and location constraints force caregivers to spend time moving between locations, rather than spending time serving their customers, which reduces throughput.

Mobile solutions that allow data entry and the retrieval of patient information to and from the electronic medical record (EMR) can reduce the need to visit the office. Data can be charted in the field rather than using a computer at the office, thereby eliminating a step in the process. This would be true if data entry were the only reason why caregivers visited the office. There may, however, be other time and location constraints. In this case, the caregivers also have to collect and return the customers' home keys. This causes another time and location constraint that offsets the potential of the mobile data entry and retrieval system. There are, however, technologies that allow a mobile platform to turn into a door-opening device, combining two technologies that together may have the power to break the constraint.

The constraints matrix (Figure 2) is an instrument that helps visualize the relationship between technologies and constraints, once these are identified. Figure 2 shows an example of the matrix based on the observations from the case study, including the previously presented time and location constraints that constitute major everyday obstacles to a swift even flow [20]. There are other reasons for visiting the office as well, such as team meetings or collecting and sorting customers' medication. These activities, however, differ from the constraints used in the example in that they do not have to be performed every day, and can be scheduled.

The technologies are wireless mobile platform-based solutions. The technologies have different capabilities, each breaking some of the constraints. Technology 1 affects constraints 1, 2, and 4, while technology 3 only affects constraint 1. The technologies considered for implementation included software applications which make possible either one-way or two-way charting of the EMR, i.e., wireless data entry, or both wireless data entry and retrieval, as well as flexible mobile scheduling, i.e., mobile work lists. One technology was a Bluetooth-based solution for wireless door-opening. The solution can be integrated into any Bluetooth-equipped mobile device. The capabilities of the technologies are:

- Technology 1: two-way charting and scheduling
- Technology 2: two-way charting and scheduling
- Technology 3: one-way charting (data entry)
- Technology 4: Bluetooth-based wireless door-opening

Not all constraints have the same impact on the system. Some are more severe than others. It is likely that the home key constraint will have a lesser impact on the system's throughput than the mobile data entry and retrieval technology. The caregivers' information processing activities consume considerably more time than customers' home keys logistics. However, due to synergy advantage, together the technologies are bound to contribute more than the sum of their parts, as they can remove an entire process step.

Constraints

Technology Solutions	1) Data Entry	2) Data Retrieval	3) Keys	4) Work Lists
Technology 1	x	x		x
Technology 2	x	x		x
Technology 3	x			
Technology 4			x	

Figure 2. Constraints matrix showing which constraints the alternative technologies affect.

Unfortunately, not all technologies are equally easy to implement. The technology, and the ease or difficulty of implementation, is termed feasibility. It incorporates the following variables: a technology's usability; the cost of the investment; the process changes involved, and their ease of implementation. There may be considerable differences both in terms of operational feasibility and in the amount of education and training required. Technologies by themselves rarely have the capability to increase throughput and improve productivity. Their power lies in their ability to allow the redesigning of the way in which tasks are performed, so that more can be done with less.

For example, mobile data entry and retrieval technologies can allow home care caregivers to spend more value-adding time with their customers, improve the quality of the EMR information as a result of timely data entry, and remove the need for double data entry. Most activities will still need to be performed, only now at a more desirable time and place, which in turn may improve the operational flow and process throughput (quality improvement), or the same amount of customers can be treated with a smaller workforce (increased productivity).

When evaluating technologies it is imperative that the process redesign be accounted for. Thus, before evaluating new technologies on the basis of financial measures such as ROI and payback time, the ease of implementing the accompanying process changes should be evaluated. Even if the time and location constraints which currently force the caregivers to visit the office were to be removed, will the caregivers stop going there? There might be other reasons for visiting the office, such as social purposes. By implementing certain technologies the constraints can be eliminated, but the true benefits will not be realized before the operating procedures and practices change.

Goldratt et al. [31] explain that constraints force organizations to create rules (policies or routines) to cope with or work around exisiting constraints. Apparently, it is not uncommon for these rules to remain after the constraint has been removed. Failure to revise the rules has been shown to keep organizations from realizing the benefits of new technologies. The authors believe routine changes can be brought about by implementing suitable incentives.

B. *Constraints vs. Non-Constraints: Timing Matters*

The quantitative analysis of operational data showed the existence of considerable peaks in the workload. Figure 3 shows the fluctuation in the workload throughout the day, in terms of the total amount of service time. Roughly fifty percent of the services are performed during peak hours, from 8-11 a.m. This causes a peak time resource constraint during the morning rush hour, throughout which the system has trouble coping with demand.

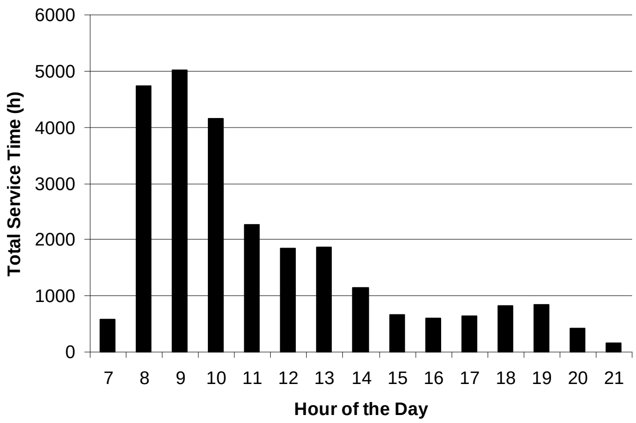

Figure 3. The caregivers' workload distribution. The figure shows the accumulated service time per hour of the day over a six-week period.

The intended mobile technology implementation sought to improve productivity by saving time on office tasks. However, the office tasks are performed in the afternoon during a period of low demand. Because of the uneven workload distribution, the time would be saved during hours of excess capacity (i.e., a non-constraint). TOC tells us that throughput can only be increased by alleviating or breaking the constraint, not by increasing the efficiency of non-constraints. In practice, saving time in the afternoon would not allow the organization to serve more customers. Consequently, implementing the solution in the current system is unlikely to save time during the peak time constraint.

By examining the home care operations through the TOC lens, we concluded that the sought technology implementation would not have the desired affect on productivity. The analysis also gave us the insight that the workload distribution was constraining the productivity of the system. However, once the workload is more level, alleviating the time and location constraints by implementing the technology is disposed to enable the organization to service more customers, improving productivity.

VI. Discussion

Assessing the technologies from a traditional cost perspective – comparing the costs of the investment with its potential time savings multiplied by cost per hour – would not have provided a true representation of the effects on the system. First, reducing the required caregiver time in the afternoon does not reduce the organization's operating expenses, since labor is a fixed cost. The capacity remains unaffected, but capacity utilization is reduced, which spreads the fixed costs over less service time. Second, an investment in technology increases operating expenses. Third, in the current mode of operation, as *time saved in the afternoon cannot be used to treat more customers* (there is already overcapacity in the afternoon), no additional throughput is gained. Consequently, there is no added throughput to offset the increase in operating expenses. This reduces productivity, the complete opposite of what the planned investment sought to achieve.

A. *Limitations of the Study*

As a result of a lack of accurate data on the timing and duration of office tasks – unlike customer encounters, back-office activities, such as office tasks and transit, are not recorded – the contribution of office activities to the total workload could not be accurately quantified.

The capacity utilization rate (CUR) of the caregivers would have been a more appropriate measure than throughput for analyzing the workload. The distribution of throughput (i.e., service time) was therefore used as an approximation of the workload. From a previous activity-based costing investigation it was known that the office tasks consume less than 20 percent of the caregivers' total working time. Transit accounts for approximately 12 percent, and is naturally distributed alongside the throughput. At the time of the study, the ratio of service time to back-office time was less than forty percent. From this it can be construed that currently there is excess capacity in the afternoon, even though it is not quite as remarkable as Figure 3 would suggest, as the figure does not account for back-office activities.

Because of the problem-solving nature of the investigation, only one organization was studied. Both the literature [28] and the authors' experience, however, suggest that the morning demand peak is a common but poorly understood characteristic of home care operations.

B. *A Note on Financial Analysis*

When constraints and their priorities are understood, organizations can proceed to financial evaluation. Financial evaluation methodology has been widely discussed in the literature [32][33][34][35][36] and basic concepts can be applied to healthcare technology evaluations. We briefly present two commonly used methods, cost-benefit analysis and cost-effectiveness analysis, as examples of basic financial evaluation tools. Cost-benefit analysis (CBA) uses a monetary frame of reference to evaluate both outcomes and costs. Both the costs of interventions and the values of outcomes are assessed in terms of money. This analysis is particularly useful if the outcomes exceed costs and the solution with the largest net benefit (outcomes subtracted for costs) should be selected. Unlike CBA, the focus of cost-effectiveness analysis (CEA) is on the non-monetary outcomes of an intervention, such as a health improvement. CEA compares the cost of alternative (intervention) outcomes. "Alternatives are calculated and presented in a ratio of incremental cost to incremental effect" [37]. CEA is therefore more suitable for assessing technologies that aspire to improve health outcomes than for evaluating process technologies.

Depending on the definition of throughput (money, visits, customers, service time etc.), these methods can be used to evaluate the financial implications of a change in throughput. The focus should, however, be on the throughput of the primary constraint, not the non-constraints'. Following TA, technologies could be assessed on the basis of their ability to increase the measure 'throughput per constraint minute' [3].

C. *When is the TOC Approach Appropriate?*

Prioritizing throughput over cost is particularly suitable for organizations that have sufficient demand to absorb the increased throughput. If, however, demand is a constraint (market constraint [12]), organizations should shift the focus to producing the same services with fewer resources, i.e., maintaining throughput while reducing operational expenses. Subject to a market constraint, the bottleneck resource that governs throughput, to which the rest of the system should be subordinated, is the resource with the highest utilization after market demand is satisfied [38]. As such, the TOC approach can help improve productivity even in the absence of an actual resource constraint.

The framework provides a rough, easy, and simple approach to ranking alternative technologies, according to their ability to affect the true operational limitations, i.e., the constraints. It is advisable to use this framework as a first step in the right direction. The suggested approach stresses that increasing throughput will reduce costs, while reducing costs will ultimately reduce throughput. Therefore the primary focus of technology assessment should be on the ability of technologies to increase throughput rather than reduce costs, so that existing resources can be used to their full potential.

VII. Conclusion and Future Work

New technologies are traditionally assessed on the basis of simplified financial estimates, with the operational impact being overlooked. This paper suggests an alternative approach based on Operations Management (OM), following the theory of constraints (TOC). The TOC approach is more pragmatic and suitable for technology evaluation in a fast-paced and growing health technology market. It holds that the performance of the system can only be increased by improving the performance of the primary constraint. Therefore technologies should be assessed on the basis of their ability to improve the performance of the constraint. The suggested approach further stresses that prioritizing throughput over cost will reduce costs, while focusing on cost reduction will ultimately reduce throughput. Once the operational aspects have been assessed, financial methods are bound to be more reliable and useful.

The empirical study showed that the time and location constraints were not the organization's primary constraint, as originally conceived. Alleviating or removing these constraints would therefore not have the desired affect. This provided the insight not to invest in the planned mobile solution, under the current mode of operation. The analysis further prompted the need to alleviate the impact of the peak time workload in Espoo Home Care by leveling service provision. An investigation of the causes behind the uneven workload distribution is currently taking place.

Suggested future work includes further development of the methodology and empirical testing of the presented framework in other settings, particularly those in which a technology is actually implemented. Comparing the use of this instrument to evaluations performed with other methods would be a valuable contribution.

Further testing and reporting on the suitability of the suggested framework for evaluating investment decisions outside healthcare is encouraged. Although TOC aspires to be the basis for all decision making, there seems to be a lack of published empirical studies on using it for this type of investment analysis.

Author Contributions

J.G., K.R, and P.L. were responsible for the conception of the study and drafted the manuscript. J.G. performed the empirical investigation, gathering and analyzing the data.

Acknowledgments

The authors wish to express their gratitude to the staff of Espoo Home Care, who were instrumental in making the study possible. In addition, Teemu Mustonen of Jyväskylä University and Ecomond Ltd. deserves special recognition for assisting in processing the quantitative raw data. The authors would further like to thank Professor Erkki Vauramo and Dr. Anu Helkkula of Aalto University, as well as the anonymous reviewers, for their valuable comments, which greatly helped to improve this paper.

During the preparation of the manuscript, J.G. and K.R. were visiting researchers at Stanford University (California). J.G. was affiliated with the Mechanical Engineering Design Group and K. R. was affiliated with the Stanford Prevention Research Center (SPRC). Both departments deserve acknowledgment for providing a stimulating research enviroment.

This study (PARETO) was funded by the European Regional Development Fund (ERDF).

References

[1] P. J. Groop, K. H. Reijonsaari, and P. M. Lillrank, "A Theory of Constraints Approach to Health Technology Assessment," in *eTelemed 2010, Second International Conference on eHealth, Telemedicine, and Social Medicine*, St. Maarten, Netherlands Antilles, pp. 147-152, February 10-16, 2010.

[2] J. Concato, N. Shah, and R. I. Horwitz, "Randomized, Controlled Trials, Observational Studies, and the Hierarchy of Research Designs," *The New England Journal of Medicine*, vol. 342, no. 25, pp. 1887-1892, Jun. 2000.

[3] E. Goldratt, *What is this thing called theory of constraints and how should it be implemented?* Croton-on-Hudson, NY: North River Press, 1990.

[4] M. Gupta and L. Boyd, "Theory of constraints: A theory for operations management," *International Journal of Operations and Production Management*, vol. 28, no. 10, pp. 991-1012, 2008.

[5] A. M. Breen, T. Burton-Houle, and D. C. Aron, "Applying the Theory of Constraints in Health Care: Part 1-The Philosophy.," *Quality Management in Health Care*, vol. 10, no. 3, p. 40, 2002.

[6] M. G. Hunink, "In search of tools to aid logical thinking and communicating about medical decision making," *Medical Decision Making*, vol. 21, no. 4, p. 267, 2001.

[7] R. A. McNutt and M. C. Odwazny, "The theory of constraints and medical error: a conversation with Robert A. McNutt," *Quality Management in Healthcare*, vol. 13, no. 3, p. 183, 2004.

[8] J. Schaefers, J. Colin, R. Aggoune, and M. Kucina, "A contribution to performance measurement in the healthcare industry: the industrial point of view," *International Journal of Business Performance Management*, vol. 9, no. 2, pp. 226–239, 2007.

[9] S. Sadat, Theory of constraints for publicly funded health systems. Dissertation: University of Toronto (Canada), 2009.

[10] M. Gupta and J. Kline, "Managing a community mental health agency: a theory of constraints based framework," *Total Quality Management & Business Excellence*, vol. 19, no. 3, pp. 281–294, 2008.

[11] J. Motwani, D. Klein, and R. Harowitz, "The Theory of Constraints in Services: Part 2 - Examples from Health Care," *Managing Service Quality*, vol. 6, no. 2, pp. 30 - 34, 1996.

[12] B. Ronen, J. S. Pliskin, and S. Pass, *Focused Operations Management for Health Services Organizations*. San Francisco: Jossey-Bass, 2006.

[13] J. Wright and R. King, *We All Fall Down: Goldratt's Theory of Constraints for Healthcare Systems*. Great Barrington, MA: North River Press, 2006.

[14] J. E. Wennberg, *Tracking Medicine: A Researcher's Quest to Understand Health Care*. New York: Oxford University Press, 2010.

[15] E. Goldratt and J. Cox, *The Goal*. Croton-on-Hudson, NY: North River Press, 1984.

[16] S. U. Rahman, "Theory of constraints: a review of the philosophy and its applications," *International Journal of Operations and Production Management*, vol. 18, pp. 336-355, 1998.

[17] J. Motwani, D. Klein, and R. Harowitz, "The Theory of Constraints in Services: Part 1 - The Basics," *Managing Service Quality*, vol. 6, pp. 53–56, 1996.

[18] J. Kujala, P. Lillrank, V. Kronström, and A. Peltokorpi, "Time-Based Management of Patient Processes," *Journal of Health Organization and Management*, vol. 20, no. 6, pp. 512–524, 2006.

[19] J. H. Blackstone, "Theory of Constraints - A Status Report.," *International Journal of Production Research*, vol. 39, no. 6, pp. 1053-1080, Apr. 2001.

[20] R. Schmenner and M. Swink, "On theory in operations management," *Journal of Operations Management*, vol. 17, no. 1, pp. 97-113, 1998.

[21] R. Schmenner, "Manufacturing, service, and their integration: Some history and theory," *International Journal of Operations and Production Management*, vol. 29, no. 5, pp. 431-443, 2009.

[22] D. Jacob, S. Bergland, and J. Cox, *Velocity: Combining Lean, Six Sigma and the Theory of Constraints to Achieve Breakthrough Performance-A Business Novel*. New York: Free Press, 2009.

[23] A. Morton and J. Cornwell, "What's the difference between a hospital and a bottling factory?," *British Medical Journal*, vol. 339, no. 20, p. b 2727, 2009.

[24] E. Goldratt and R. E. Fox, *The Race*, 1st ed. Croton-on-Hudson, NY: North River Press, 1986.

[25] B. Ronen and Y. Spector, "Managing system constraints: a cost/utilization approach," *International Journal of Production Research*, vol. 30, no. 9, pp. 2045–2061, 1992.

[26] L. Boyd and M. Gupta, "Constraints management: What is the theory?," *International Journal of Operations & Production Management*, vol. 24, no. 4, pp. 350 - 371, 2004.

[27] E. Goldratt, *The haystack syndrome: Sifting information out of the data ocean*. North River Press Great Barrington, MA, 1990.

[28] P. Eveborn, P. Flisberg, and M. Rönnqvist, "Home Care Operations," *OR/MS Today*, vol. 31, no. 2, pp. 38–43, 2004.

[29] S. Agnihothri, N. Sivasubramaniam, and D. Simmons, "Leveraging technology to improve field service," *International Journal of Service Industry Management*, vol. 13, no. 1, pp. 47–68, 2002.

[30] K. H. Ilvonen, P. J. Groop, and P. M. Lillrank, "Mobile Services Provide Value by Decoupling the Time and Location Constraints in Healthcare Delivery," in *eTelemed 2009, First International Conference on eHealth, Telemedicine, and Social Medicine*, Cancun, Mexico, pp. 216-219, February 1-7, 2009.

[31] E. Goldratt, E. Schragenheim, and C. A. Ptak, *Necessary But Not Sufficient*. Croton-on-Hudson, NY: North River Press, 2000.

[32] T. T. Edejer et al., "WHO guide to cost-effectiveness analysis," *WHO, Geneva*, 2003.

[33] M. Johannesson and B. Jönsson, "Economic evaluation in health care: is there a role for cost-benefit analysis?" *Health Policy*, vol. 17, no. 1, p. 1, 1991.

[34] H. E. Klarman, "Application of cost-benefit analysis to health systems technology," *Journal of Occupational and Environmental Medicine*, vol. 16, no. 3, p. 172, 1974.

[35] K. E. Warner and R. C. Hutton, "Cost-benefit and cost-effectiveness analysis in health care: Growth and composition of the literature," *Medical Care*, vol. 18, no. 11, pp. 1069–1084, 1980.

[36] A. E. Boardman, D. H. Greenberg, A. R. Vining, and D. L. Weimer, Cost-benefit analysis: concepts and practice, 3rd Ed. Upper Saddle River, NJ: Prentice Hall 2006.

[37] J. Brazier, J. Ratcliffe, and A. Tsuchiya, *Measuring and valuing health benefits for economic evaluation*. New York: Oxford University Press, 2007.

[38] J. Balakrishnan, C. H. Cheng, and D. Trietsch, "The Theory of Constraints in Academia: Its Evolution, Influence, Controversies, and Lessons," *Operations Management Education Review*, vol. 2, pp. 97–114, 2008.