Reflections
VOLUME 2 | ISSUE 1 | 2018
ON HEALTHCARE MANAGEMENT

WHAT IS CRITICAL CHAIN PROJECT MANAGEMENT AND HOW CAN IT BE USEFUL IN HEALTHCARE?

PAUL WILKENS, MBA

ABSTRACT

This paper reviews the basics of Goldratt's Theory of Constraints (TOC), followed by a discussion of how they have been adapted for project management (PM). Some of the benefits and flaws of this management method are described, followed by a literature review of known applications to healthcare organizations, which concludes with a summary of healthcare-specific observations. *Note:* Project management based on the Theory of Constraints is described variously as Critical Chain Project Management (CCPM), Critical Chain / Buffer Management (CC/BM), Critical Chain Scheduling / Buffer Management (CCS/BM), TOC-PM, and many others. This review uses “CCPM” throughout.

KEYWORDS

theory of constraints, operations management, project management, critical chain

AUTHOR BIOGRAPHY

Paul Wilkens is a 2017 graduate of The Healthcare MBA. You can reach him at: Division of Management, Oregon Health and Science University, Mailcode L473; Portland, OR 97239, or wilkens@sbcglobal.net.

doi: 10.6083/c4xe-za76

Introduction

After twenty years of constant development, the Toyota Production System (TPS) became a benchmark for organizing industrial output to consistently achieve the highest levels of efficiency, product quality, and throughput. The original TPS was highly customized for the factory floor, but the principles behind TPS can be generalized, analogized, and broadly applied to many different types of production, including healthcare. TPS practices that focus on efficiency contributed to the development of LEAN methodology. TPS practices that focus on product quality contributed to the development of Six Sigma. TPS practices that focus on throughput contributed to the development of the Theory of Constraints.

The Theory of Constraints on the Factory Floor

Goldratt observed that the maximum throughput of a factory system could never exceed the throughput of the slowest processing unit. In other words, total system throughput is constrained by its slowest atomic process, which becomes the choke point or "weakest link." Goldratt's *The Goal* is a business novel that describes several different methods for finding and maximizing the throughput of choke points [Goldratt & Cox, 2004]. After numerous attempts and partial failures, Goldratt's factory manager develops a version of the TPS drum-buffer-rope system. The rope is the customer order, and its due date "pulls" the product through the factory. The drum is controlled by materials release, which prevents workers from doing too-early work on non-urgent items during downtime in hopes of superficially improving their efficiency ratings. The buffer is large enough to ensure that the slowest processing unit always has work waiting despite the natural variability introduced by occasional mishaps and delays in prior production lines, but small enough to prevent waiting product from accumulating and interfering with floor operations.

The Goal provides an imaginative high-level description of the throughput-enhancing aspects of TPS, which in its original form is so highly customized for the production of automobiles that it can be difficult to generalize. *The Goal* is not overly simplistic in its description. The factory manager becomes the modern equivalent of a philosopher-king; at one point, he pulls his college text on Greek philosophy off the shelf to help him analyze a particularly vexing production flow problem. He is surrounded by dedicated, agile, and competent staff. They spend most of their time in a constant search for new bottlenecks which were created by unintended consequences of previous improvements. There are no easy answers, and no set formulae to follow. At the time, it contradicted some detrimental fallacies in traditional factory management by describing how a factory with all workers spending 100% of their time processing parts – thereby achieving perfect efficiency – could still miss production dates and cost targets. Retooling to fulfill nonessential orders seems productive but can interfere with critical orders. Efficiency measurements can do more harm than good when poorly-chosen metrics encourage counterproductive activity.

There are also flaws in Goldratt's narrative. In *The Goal* throughput-enhancing innovations improve factory morale so much that sheer pride becomes the only form of quality assurance needed, but in reality, most factories still depend on Six Sigma methods for quality maintenance. The methods they use to improve the throughput of constrained resources were borrowed from TPS, but are today best described as LEAN. By concentrating on throughput, he artificially diminished the importance of quality and efficiency, the other two legs of the tripod.

Goldratt's Theory of Constraints relies on five steps:

1. Identify the system's constraint.
2. Decide how to exploit it.
3. Subordinate and synchronize everything to these decisions.
4. Elevate the system's constraint.
5. When the constraint moves, return to step 1.

Only one resource can be busy all of the time. "If a system is performing as well as it can, not more than one of its component parts will be. If all parts are performing as well as they can, the system as a whole will not be." [Leach, 2005, p. 45] There will always be a bottleneck. Where is the best place for it?

Applying the Theory of Constraints to Project Management

In mass production environments, each individual component part has a low "touch time" and a high "wait time." Product components spend a lot of time waiting to be processed by a machine, but very little time in the machine, and ratios higher than 100:1 are common [Goldratt & Cox, 2004]. Products that are crafted manually are exactly the opposite. Manual crafting requires high touch times. With few exceptions, facilities for treating patients are more similar to crafting workshops than to mass-production factories. Each patient spends significant time with staff and providers, and the ratio of touch time to wait time is closer to 1:1. Project management is also more similar to craftwork than to mass production. The most constrained resources are often people, who spend many hours a day working on specific project tasks. Because of these profound underlying similarities, *any successful application of TOC to healthcare environments will necessarily resemble the application of TOC to project management environments.*

Can we adapt ideas taken from mass production and successfully apply them to such different environments? In a later business novel called "*Critical Chain*," Goldratt (1997) made several modifications to his previous Theory of Constraints, and applied the concepts to project management. His recommendations are based on a careful and philosophical consideration of principles he thought particularly important to project management. He believed that participants usually overestimate the amount of time needed to complete each task, and then work expands to fill the available time, or workers reduce their effort to take advantage of the generous estimate, or both. Goldratt believed that for individual workers, multitasking between projects is about 25% less efficient than constant work on a single project. He believed that traditional project management runs into frequent conflicts, especially in a multi-project environment, because it does not sufficiently account for the availability of constrained resources. He identified counterproductive social forces in project management, including the tendency to overtask the best workers, and the tendency to start projects too early just because a contract was signed and everyone wants to look busy.

He formulated an alternative system of individual project management, with these key elements:

- Assume that the estimate for each task contains a 50% safety factor. Remove this by reducing each task length by half, then add half of the remainder to the end of the project. This is the project buffer. If your project's task path was 80 days long, it is now 40 days long, with a 20-day safety buffer at the end.
- Arrange the baseline schedule's use of constrained resources with respect to precedence and availability. The "critical chain" is the set of tasks that utilize the most constrained resources. Run tasks in parallel to reduce total time and work in progress. Forbid multitasking.
- For task sequences that feed into the critical chain, but are not themselves part of the critical chain, establish "feeding buffers" to ensure that they complete before the constrained resource becomes available, thereby keeping the constrained resource busy.
- Working backwards from the delivery date, use the buffered timeline to establish firm start dates for all tasks on the critical chain. Eliminate the feeding buffers from the non-critical tasks, and ask workers to complete everything as quickly as possible, regardless of schedule. Monitor buffers for warning signs, but ignore individual due dates for buffered tasks as long as the buffer seems to

be sufficient. Natural variability is expected and will not affect the final project completion date. Use resource buffers (notifications of impending tasks) to warn constrained resources of upcoming expectations.
- Continually evaluate and recalculate buffer use, the critical chain pathway, and the scheduling of constrained resources, looking for ways to complete the project earlier than planned. Be alert for adjustments to constrained resources that alleviate the constriction to such an extent that other elements become new choke points.

He also briefly describes how to schedule multiple projects that require access to the same constrained resources. First, prioritize and plan the individual projects, as above. When a project needs access to a constrained resource, place a capacity buffer before the requirement to ensure that the feeder tasks will be completed before the resource is available, ensuring that the resource is always busy. Continually measure and manage the buffers.

The primary objective of Critical Chain Project Management (CCPM) is to "create a precedence and resource-feasible baseline schedule that minimizes the project duration" [Herroelen, Leus, & Demeulemeester, 2002, p. 50]. Keeping projects short has at least three advantages. First, cash flow is improved. Second, costs related to delays and contingencies are reduced. Third, the customer has less time to change the scope of the project [Steyn, 2002].

Advantages and Disadvantages of CCPM

These methods are highly optimized for project organizations. A sincere attempt to implement CCPM will require considerable thought and discussion on topics of critical importance to the success of projects. How do we estimate durations? Which are the key tasks? What are the key resources? When are they needed? What resources are most highly constrained? How should we determine precedence when two tasks or projects need the same constrained resource?

Regardless of project management methodology, the act of considering these factors is likely to promote project success. CCPM also emphasizes constant reevaluation and reassessment, which is generally helpful to the practice of project management.

By aggregating the effects of uncertainty and natural variability, CCPM smooths statistical fluctuations. By adjusting the links in the chain so that everything moves at roughly the same speed, CCPM allows organizations to allocate resources more efficiently while maintaining throughput.

Since Goldratt's initial description, studies have also identified various problems with CCPM:

- The assumption that 50% of task time is safety buffer is often untrue. CCPM emphasizes that due dates are unimportant, which should prevent workers from further buffering their estimates, but in reality, the way estimators react to this new method is unpredictable and potentially counterproductive. If there is a way to reliably estimate the true length of a task, it would be better to start with that and add buffer at the end, rather than extracting the buffer from individual tasks with uncertain and variable safety margins [Raz, Barnes, & Dvir, 2004].
- There is no evidence for the twin notions that work expands to fill time available and workers work only as fast as the schedule requires. One major study of software development projects found that 60% of tasks were completed early, 8% were completed on time, and 32% were late, when compared with initial estimates by the project workers [Hill, Thomas, & Allen, 2000].
- The usefulness of the project buffer is algorithm-dependent. The root-square-error method is demonstrably superior to the 50%

assumption for longer project task sequences. The most advanced methods attempt to account for the destabilizing effects of interdependence[Bie, Cui, & Zhang, 2012].

- CCPM techniques work best when critical chain tasks are not in sequence. With significant serialism, the critical chain becomes equivalent to the "leveled" critical path, and little is gained [Herroelen & Leus, 2004].
- There are often many different ways to draw the project network [Herroelen, Leus, & Demeulemeester, 2002]. CCPM is optimized for manufacturing tasks, like construction and assembly. With more complex projects that contain multiple interdependencies, there are many different ways to construct the timeline, and it is not always obvious where feeding buffers are most needed. Buffers also significantly increase the number of items on the Gantt chart, making it harder to read [Herroelen & Leus, 2004].
- Feeding buffers are often underestimated, and feeding tasks can quickly become part of the critical chain, which can alter task prioritization and project timelines [Herroelen & Leus, 2004].
- Milestones can serve important social purposes like enhancing morale and promoting group cohesion [Updegrove, 2014]. Eliminating them may have undesirable side effects.
- Rescheduling of tasks based solely on CCPM measurements can be insensitive to other factors like contract bonuses and penalties. In complex projects, significant rescheduling of non-critical chain tasks will often affect critical task start dates, which are not supposed to change [Herroelen, Leus, & Demeulemeester, 2002; Herroelen & Leus, 2004].
- The harmfulness of multitasking is highly dependent on both the type of task and the overall context. Multitasking does not always reduce efficiency by 25%; in multi-project environments with low variability, it can be both useful and harmless [Herroelen & Leus, 2004].

According to Raz *et al*: "Proponents of CCPM have claimed some dramatic successes, although from our personal experience these appear to be mainly in organizations that started out with weak or non-existent project management methodologies" [Raz, Barnes, & Dvir, 2004, PDF page 15 (page number will vary by format and access method)]. This observation is particularly relevant to healthcare providers, because most do not utilize project management.

Literature Review

Healthcare organizations are more likely to utilize the LEAN and Six Sigma aspects of TPS, and relatively few case studies of TOC and CCPM have been published. Because project management and healthcare provision are fundamentally similar – both rely on highly-trained experts who execute tasks with high touch and low wait times – we will review the application of both CCPM and TOC to healthcare organizations.

Umble & Umble (2006)

Background: CCPM techniques were used to improve waiting times for surgery in Britain's NHS, where lead times are unusually long. Staff attempted to reduce processing to under four hours, and total check-in time to under twelve hours.

Method: Experienced implementers created a color-coded display that helps staff expedite patients who are waiting too long. During processing, patients are in the green buffer for the first hour, the amber buffer during the second hour, and the red buffer during the third and fourth hours. During check-in, patients go from green to amber after two hours and from amber to red after four hours. Staff concentrate on expediting patients in the red buffer. Staff also document the reasons for delay, which generates helpful long-term statistics. No other changes were made.

Results: At both hospitals, the number of patients processed in under four hours was between 50% and 70%. After implementation, rates exceeded 90%. Check-in improvements were similar. The results were durable.

Discussion: Project management is often non-existent in provider settings, so even limited efforts can have outsized effects. In this case, the simplest possible intervention led to substantial improvement, at little cost. We might expect further refinements to land on a higher point of the curve of diminishing returns.

Lubitsh, Doyle, & Valentine (2005)
Background: TOC techniques were used to reduce waiting list times for three surgical departments managed by Britain's NHS.

Method: The effort began with a TOC workshop at each department, where participants tried to identify their system's constraint. In ENT, the constraint was inpatient nurses. In Eyes, the constraint was operating room availability, but further investigation showed that OR teams were spending 30% to 40% of their time waiting for feeding tasks. Interventions were customized to specific department needs, and numerous quantitative metrics were collected over 40 months.

Results: Several metrics improved significantly in the ENT and Eyes departments. Neurosurgery did not improve.

Discussion: The authors attribute the results to differences in variability and interdependency. The TOC methods were much easier to implement in the Eyes department, a silo operation with little outside dependency. Neurosurgery patients presented with greater variability and required complex preparation. The department was much more interdependent with other departments, causing more variability. Neurosurgery staff did not participate in the TOC process as actively as the other two. The authors also found that "local customization" was particularly "important to acceptance," which suggests that participants were very critical of generic recommendations. This is consistent with Goldratt's requirement for careful study of specific context, and his expectation that methods must be continually adjusted as constraints move around the system.

Gupta & Kline (2008)
Background: The authors used TOC to improve patient throughput at a community chemical dependency clinic.

Method: After identifying intake processing and psychiatric evaluations as the two most constrained throughput steps, the authors made several improvements in both areas, extending and increasing pre-appointment contacts with patients and redesigning the office layout to increase throughput by reducing inefficiencies.

Results: The "no show" and cancellation rates fell from 43% to 20%. Compliance with medication did not improve.

Discussion: The authors identified two slow processes, studied them closely, and made internal improvements. As a result, more patients were seen, fewer missed their appointments, and appointments took less time. While a LEAN approach would have produced nearly the same immediate results, the authors made an additional effort to identify parts of the system that were most likely to become the next constrained resource as improvements to the initial constraints shifted burdens elsewhere. For instance, the improvements in clinical throughput increased the paperwork to such an extent that clerical staff could no longer complete their daily tasks during normal working hours. Accurate predictions of unintended consequences are rare in the literature.

Kershaw (2000)
Background: The authors identified bottlenecks and improved patient throughput at an outpatient oncology clinic.

Method: For every step of the treatment process, resource availability was measured and compared with patient needs. There were too few treatment chairs, so the authors attempted to elevate that constraint, and increase efficiency. IV access was established ahead of time, patient education was combined with drug administration, medication order routing was improved, and chairside supplies were made more accessible.

Results: Capacity increased by 25 to 30 patients per day. Average treatment time decreased from 2.5 to less than 2 hours.

Discussion: The authors first tried changes that required more staff time, but the increased work damaged morale. Other than the first step of identifying constraints, this effort was very similar to the LEAN approach.

Womack & Flowers (1999)
Background: The authors used TOC to shorten primary care wait times and primary care appointment times for the USAF 366th Medical Group.

Method: Previous cost-cutting and efficiency initiatives made the medical technician the most constrained resource, not the primary care provider. They made several changes to improve efficiency in both the scheduling process and the processes surrounding the patient-physician visit.

Results: Average wait times decreased from 17 days to 4.5 days. Appointment duration decreased from 20 minutes to 15 minutes with no reduction of patient-physician contact time.

Conclusion: The authors also determined that they could expand the patient base by 800, resulting in $1.6 million of additional revenue, for an additional cost of $200,000. This study is nearly two decades old, but more recent studies indicate that there are still plenty of badly-needed, low-effort, large-impact projects at most healthcare provider organizations.

Several other papers have been written on the application of various aspects of CCPM and TOC to healthcare organizations. Rotstein, Wilf-Miron, Lavi, Seidman, Shahaf, Shahar, Gabay, & Noy (2002) calculated the cost-effectiveness of hiring a new ED physician and found that the physician was the most constrained resource only when patient volumes are high. Sadat, Carter, & Golden (2013) developed a discrete event simulation model that implements drum-buffer-rope concepts to explore maximization tradeoffs between two scenarios: "low wait times" and "high constraint utilization”. Wolstenholme (1999) developed a model to demonstrate that providing additional post-hospitalization "intermediate care" would improve throughput more than adding hospital beds to the NHS system. Young, Brailsford, Connell, Davies, Harper, & Klein (2004) discuss the general applicability to patient care of techniques developed for industrial systems, including LEAN, Six Sigma, and TOC. Breen, Burton-Houle, & Aron (2002) review prior applications of TOC to healthcare organizations, and make recommendations for successful use of CCPM.

Conclusion

Critical Chain Project Management's usefulness to healthcare organizations is affected by several unique factors. First, while improved patient throughput is clearly beneficial in many different environments, there’s no obvious reason to conclude that maximized patient throughput will lead to maximized population health. The ultimate goal of the healthcare system is to promote health, not throughput, and the two are most closely related during disease treatment, not prevention. Second, identifying bottlenecks can be particularly difficult in healthcare settings. For maximum efficiency and return on investment, the providers with the rarest licenses should be the most constrained resources, and all other processes should keep them working at 100% utilization. In fact, many of these studies have found that the physicians are not the most constrained, and are often not operating at the top of their license. Finally, most healthcare providers have so completely failed to utilize even the most basic project management techniques that an advanced technique is not yet needed; the low-hanging fruit can be plucked with minimal guidance from CCPM, or LEAN, or Six Sigma, or TOC, or Agile, or with the traditional project management techniques described in standard textbooks. However, the techniques introduced by TOC and CCPM are valuable because they provide focusing questions that force participants to carefully consider resource availability, process interdependence, project precedence, and the relationship of throughput to other measures of success. Greater awareness of these issues can lead to many different kinds of improvement to healthcare organizations, and whether a TOC project management purist would consider the outcomes to be "true CCPM" is unimportant.

References

Bie, L., Cui, N., & Zhang, X. (2012). Buffer sizing approach with dependence assumption between activities in critical chain scheduling. *International Journal of Production Research*, *50*(24), 7343-7356.

Breen, A. M., Burton-Houle, T., & Aron, D. C. (2002). Applying the theory of constraints in health care: Part 1—the philosophy. *Quality management in healthcare*, *10*(3), 40-46.

Goldratt, E. M. (1997). *Critical chain: A business novel*. Great Barrington, MA: North River Press.

Goldratt, E. M., & Cox, J. (2004). *The goal: a process of ongoing improvement*. Great Barrington, MA: North River Press.

Gupta, M., & Kline, J. (2008). Managing a community mental health agency: A Theory of Constraints based framework. *Total Quality Management*, *19*(3), 281-294.

Herroelen, W., & Leus, R. (2004). Robust and reactive project scheduling: a review and classification of procedures. *International Journal of Production Research*, *42*(8), 1599- 1620.

Herroelen, W., Leus, R., & Demeulemeester, E. (2002). Critical chain project scheduling-Do not oversimplify. *Project Management Journal*, *33*(4), 46-60.

Hill, J., Thomas, L. C., & Allen, D. E. (2000). Experts' estimates of task durations in software development projects. *International journal of project management*, *18*(1), 13-21.

IEEE Engineering Management Review, *32*(2), 35-44.

Kershaw, R. (2000). Using TOC to “cure” healthcare problems. *Management Accounting Quaterly* *1*(3)(Spring), 1-8.

Leach, L. P. (2005). Critical Chain Project Management, 2nd Ed. Norwood, NA: Artech House.

Lubitsh, G., Doyle, C., & Valentine, J. (2005). The impact of theory of constraints (TOC) in an NHS trust. *Journal of Management Development*, *24*(2), 116-131.

Raz, T., Barnes, R., & Dvir, D. (2004). A critical look at critical chain projectmanagement.

Rotstein, Z., Wilf-Miron, R., Lavi, B., Seidman, D. S., Shahaf, P., Shahar, A., Gabay, U., & Noy, S. (2002). Management by constraints: considering patient volume when adding medical staff to the emergency department. *The Israel Medical Association Journal: IMAJ*, *4*(3), 170-173.

Sadat, S., Carter, M. W., & Golden, B. (2013). Theory of constraints for publicly funded health systems. *Health care management science*, *16*(1), 62-74.

Steyn, H. (2002). Project management applications of the theory of constraints beyond critical chain scheduling. *International Journal of Project Management*, *20*(1), 75-80.

Umble, M., & Umble, E. J. (2006). Utilizing buffer management to improve performance in a healthcare environment. *European Journal of Operational Research*, *174*(2), 1060-1075.

Updegrove, D. (2014) *The Critical Chain Implementation Handbook*. Boeing Company.

Wolstenholme, E. (1999). A patient flow perspective of UK health services: exploring the case for new" intermediate care" initiatives. *System Dynamics Review*, *15*(3), 253.

Womack, D. E., & Flowers, S. (1999). Improving system performance: A case study in the application of the theory of contraints. *Journal of Healthcare Management*, *44*(5), 397.

Young, T., Brailsford, S., Connell, C., Davies, R., Harper, P., & Klein, J. H. (2004). Using industrial processes to improve patient care. *Bmj*, *328*(7432), 162-164.