Lean beyond waste

Roemeling, Oskar-Paul

CHAPTER 1

General introduction

1.1 INTRODUCTION

The academic public was made aware of a promising production system used by Toyota through an academic paper in 1977 authored by Sugimori, Kusunoki, Cho and Uchikawa (New, 2007). Later, in 1984, intrigued by the low cost and high quality of Japanese cars, a group of American researchers from MIT started a study aimed at investigating the ‘secrets’ of the Japanese car manufacturers’ success. The study resulted, amongst other publications, in the book ‘The machine that changed the World’ (Womack *et al.*, 1990). The production approach now labeled “Lean” was introduced to the broader public (Shah and Ward, 2007). Moreover, the book initiated a stream of new research, a stream that has gained interest over the years and continues to grow to this day (Samuel *et al.*, 2015).

In their book, Womack, Jones, and Roos report on the findings of their longitudinal study at various international car manufacturers including Toyota, Honda, General Motors, and Ford. The study showed that the Japanese manufacturers, and especially Toyota, were able to outperform their Western counterparts. This was not because of external factors such as the Japanese culture or the Keiretsu (systems of banks, shareholders, and businesses), but because of a superior approach: i.e. Lean (Krafcik, 1988).

After its increasingly common application in manufacturing, and later service, environments, Lean is now also being applied in healthcare. Its popularity is reflected in the size of the Dutch 'Lean in healthcare' foundation (LidZ) that currently has over 60 member organizations, and well over 2000 individual participants. This shows the interest in Lean ideas in healthcare settings in the Netherlands, and underlines the importance and relevance of this thesis for practice. Perhaps surprisingly, given Lean's popularity, the effects of Lean applications in healthcare are still unclear.

In this thesis, we therefore focus on the effects and the application of Lean interventions in healthcare. Lean interventions are changes made to (and in) the work environment that are based upon the application of Lean principles. Currently, evidence for the long-term effectiveness of Lean performance improvements is limited, and studies often provide no more than anecdotal evidence. Inconclusive literature findings, coupled with the popularity of Lean management initiatives in healthcare environments, suggest an increased need for rigorous Lean-oriented studies (Mazzocato *et al.*, 2010).

Research on the application of Lean principles is sometimes able to show positive outcomes (e.g. Bamford *et al.*, 2015; Breslin *et al.*, 2014), yet literature reviews struggle to deliver a conclusive verdict on the long-term Lean effects (e.g. Mazzocato *et al.*, 2010; Poksinska, 2010). Especially in healthcare, the impact of applying Lean management principles is still varied. In this introductory chapter, we first provide a review of the Lean philosophy. Subsequently, we introduce the research projects and then present an outline of the thesis. At the end of this chapter, we provide an overview of the Dutch healthcare system for those unfamiliar with it.

1.2 THEORETICAL BACKGROUND

Womack *et al.* (1990) initially identified Lean as a manufacturing system that used fewer resources than the more commonly applied mass production methods. Central to Lean is the reduction of waste, activities, and parts of a process that do not add customer value (Womack and Jones, 1996). Within Lean, a distinction is made between seven types of waste: inventory, movement, defects, over-processing, over-production, transport, and waiting (Womack *et al.*, 1990). The creation of these wastes often has a very practical nature, for example, *movement* includes a worker constantly having to reach for an item needed during the job. A solution would be to relocate the item – move it closer to the worker. This simple example underlines a crucial aspect of waste, namely that it is avoidable. Not surprisingly, tackling waste usually features prominently in Lean practice (Chapter 3, this thesis).

Despite there being a plethora of studies that have investigated Lean (Bhamu and Singh Sangwan, 2014), the research field still lacks an agreed formal Lean definition (Pettersen, 2009; Samuel *et al.*, 2015). However, most definitions do indicate that waste is not the only important aspect of Lean. Most authors agree that continuous improvement is central to Lean (Pettersen, 2009). Through continuous improvement, and often with the help of Lean tools such as value stream maps or 5S, Lean strives to improve the flow of materials and/or information.

Flow can be improved by reducing waste and variability (Schmenner and Swink, 1998). Nevertheless, depending on the research focus, authors tend to stress different Lean aspects. For example, there are authors who are especially interested in continuous improvement (e.g. Bhuiyan and Baghel, 2005), and others who stress the importance of waste-related issues (e.g. Jimmerson *et al.*, 2005). When we consider the review by Mazzocatto *et al.* (2010) it seems that currently variability-related matters attract little interest and that obvious waste receives the most consideration.

The limited attention to variability in, especially empirical, Lean research is surprising given that various authors report on the importance of variability. For example, Shah and Ward (2007) stress the role of variability in Lean in their statement that Lean's main objective is to "eliminate waste by concurrently reducing or minimizing supplier, customer, and internal variability". Further, according to Treville and Antonakis (2006), Lean is meant to maximize the utilization of capacity resources whilst minimizing buffer inventories through the minimization of system variability. A more specific theory explaining the importance of reducing buffers and variability is the 'Theory of Swift, Even Flow' (Schmenner and Swink, 1998) which is used as a theoretical underpinning for Lean (Shah *et al.*, 2008; Boer *et al.*, 2014).

Hopp and Spearman (2004) identify the roles of buffers and emphasize the importance of reducing variability in Lean approaches: variability in supply and demand will decrease the overall performance of a process because it leads to buffers (Hopp, 2008). In this thesis, we are especially interested in both variability and buffers, two aspects of Lean that have had very limited attention yet should be considered central to a Lean approach (e.g. Shah and Ward, 2007). Therefore, in this thesis, we adopt the definition of Lean by Hopp and Spearman (2004) that puts buffers at its heart (see Textbox 1.1).

Textbox 1.1 Lean definition.

Definition of Lean:
"The production of goods or services is *Lean* if it is accomplished with minimal buffering costs"
(Hopp and Spearman, 2004, p. 144).

In applying Lean principles, an organization should strive to reduce variability in order to reduce buffers. In essence, this is what Hopp and Spearman (2004) identified at the Toyota Motor Company. When Toyota implemented Lean, they exchanged inventory buffers for capacity buffers. The previous inventory buffers allowed sources of variability to remain 'hidden' and, when these inventories were reduced, the sources became apparent. The issues created by variability could then be overcome through having an extended capacity. Later, through a continuous improvement process, both inventory and capacity buffers could be reduced. Figure 1.1 provides a graphical representation of this process, which essentially reflects the maturity steps in Lean development. However, these maturity steps have yet to be empirically established. In the early stages of Lean development, we would expect organizations to work on direct waste reduction. Later, with increased experience, we would expect them to focus on reducing buffers and variability.

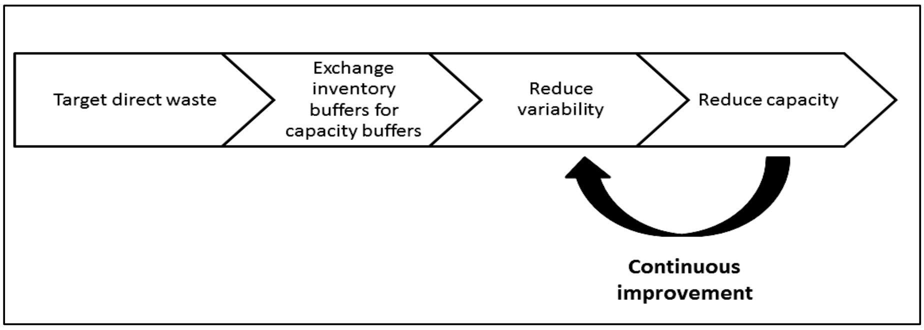

Figure 1.1. Phases of Lean adoption (Hopp, 2008, p. 91)

Variability and buffers

In healthcare, it can be helpful to differentiate between different types of variability. In this thesis, we initially distinguish between natural and artificial variabilities (Litvak and Long, 2000). Natural variability is always present, and can only be influenced or controlled to a limited extent. Conversely, artificial variability occurs because of one's own decisions and this type of variability can be addressed with the help of Lean. Kim *et al.* (2006) provides several examples that can be seen as artificial variability, such as the delayed discharge of patients because of poor planning.

Alongside natural and artificial variabilities, Litvak and Long (2000) provide an additional typology to distinguish between different types of variability. They suggest the following types of variability are relevant to healthcare: (1) flow variability related to the arrival of patients; (2) clinical variability related to the degree of patients' illnesses or differences in response to treatment; and (3) professional variability related to differences in the capabilities of medical professionals to deliver care. In this thesis, flow and professional variabilities are of prime importance since we expect clinical variability to be a purely natural affair and therefore impossible to influence. This implies that Lean interventions would have a very limited (if any) effect on clinical variability, and therefore the focus should predominantly be on reducing flow and professional variabilities.

Variability leads to buffers. It is common to distinguish between inventory, capacity, and time buffers (Hopp, 2008). However, provided healthcare is considered a service, inventory buffering is not applicable in the core healthcare process (Jack and Powers, 2004) unless one would see the patients in a process as inventory. In this thesis, we see time buffers (waiting time) as a mechanism to handle the variability in cases of waiting patients. That is, variability is buffered by having patients waiting until the required resources become available. The alternative to having patients waiting for resources, is to have idle capacity waiting for patients. In the event of having excess idle capacity, we speak of capacity buffers.

Time buffers can be used to buffer arrival variability, such as in an emergency department where non-urgent patients have to wait for treatment (Van der Vaart *et al.*, 2011). However, from a customer-oriented quality perspectives, having a waiting time for patients could be undesirable. This could mean that time buffers are not always acceptable. Conversely, capacity buffers are costly and unattractive, especially in an era of shrinking healthcare budgets. Ultimately, the only viable option seems to be to reduce the underlying variability, and this is where Lean interventions should focus once direct waste issues have been addressed.

Apart from time and capacity buffers, Hopp *et al.* (2007) also identified quality buffers as an additional buffer type that could be used in service environments. The underlying idea is that, in services, professionals can vary their processing times to cope with demand. In their study, Hopp *et al.* (2007) illustrated the principle of quality buffers in a telemarketing service. In their example, the general idea was that a telemarketer could invest additional time in attempting to sell to a client when times were quiet. The authors saw this as translating into a higher service quality. However, whether quality buffers are an appropriate extension to the established buffers in healthcare is debatable.

Lean in the context of healthcare

In this thesis, we are especially interested in the application of Lean in a healthcare context. Healthcare is considered a service, a different environment to manufacturing environments where Lean was initially introduced. The positive relationship between Lean and performance is well established in manufacturing (e.g. Seth and Gupta, 2005; Taylor, 2005; Subha and Jaisankar, 2012), and in service environments there have been many positive reports on Lean applications (e.g. Furterer and Elshennawy, 2005; Staats *et al.*, 2011; Piercy and Rich, 2009).

However, it would be dangerous to assume these positive reports can be immediately translated to healthcare. Healthcare environments have some specific features that distinguish them within the broader services spectrum. One of the main differences to most services is that the client, or part of the client, (patient or patient material) is the 'object' being worked upon. Essentially, one is working with living and breathing persons, not inanimate objects. In addition, we can think of examples where the interactions between medical professionals and clients are more varied and personal (Shah *et al.*, 2008). Overall, we can expect Lean applications in healthcare to have additional complexities related to the specific environment.

Lean is applied in the healthcare sector in response to the increased demand for higher quality care. The application of Lean in healthcare environments is a relatively new phenomenon (Burgess and Radnor, 2013) although Fillingham (2007) reported reduced lengths of stay, among other performance increases, due to Lean. In another healthcare study, Grove *et al.* (2010) showed that applying Lean tools assisted in identifying possible process improvements, such as the simplification of standard tasks and reductions in paperwork. In addition, Smith *et al.* (2012) applied various Lean tools and techniques, including Kaizen events, to improve patient safety.

These above examples, and other studies (Mazzocato, 2010), show promising outcomes. Nevertheless, there seems to be an over-focus on tools and technology, and elements such as variability are largely unaddressed (Andersen *et al.* 2014; Radnor, 2010). Furthermore, current studies are unable to indicate the long-term effects of adhering to Lean principles. In this thesis, we attempt to go beyond short-term effects, achieved through adopting Lean tools for waste reduction, and instead emphasize the roles of variability and buffers.

Joosten *et al.* (2009) claimed that a thorough understanding of the effects of Lean in healthcare was absent. Similar later observations were made by Burgess and Radnor (2013) who commented that the implementation of Lean principles is often isolated rather than organization-wide. Several authors have similarly concluded that the application of Lean in healthcare is fragmented (Young and McClean, 2008; Proudlove *et al.*, 2008; Ballé and Régnier, 2007). Curatolo *et al.* (2014) observe that most studies on Lean in healthcare lack a thorough methodological basis, and that this hinders well-founded statements concerning the Lean – performance relationship. In essence, the results of Lean research in healthcare seem promising but there is still a need for more thorough research. The current status of the research field has several implications for our own study, in the sense that our research:

- Should report on the long-term effects of applying Lean principles;
- Should go beyond the application of tools aimed at obvious waste;
- Should provide a clear understanding of the principles, beyond waste reduction, that drive Lean;
- Should provide a strong methodological basis that allows firm statements concerning the Lean – performance relationship in healthcare.

Thesis focus

In developing research projects for this thesis, we took the outstanding issues identified in the literature into account. In Chapter 2, we report on our first research project where we attempted to provide clarity on the effects of the long-term application of Lean principles in a medical setting. We investigate if performance does indeed improve, perhaps gradually, over time because of a series of Lean interventions. The results indicated that Lean interventions had only a limited impact on time buffer reduction. This led us to question the focus of practical interventions, and therefore, in our second study, we investigated a large number of interventions. Could we expect the impact of Lean interventions on variability and buffers when we consider their actual content?

In Chapter 3, we show the results of this second study and we explore whether knowledge on variability and buffers can influence the focus of Lean interventions. We concluded that the most prevalent focus is on waste, and that variability and buffers seem to be absent from the practitioners' mind-sets. Given our belief that flow-oriented Lean projects should address variability and buffers, we turned our attention to these kinds of projects in our next study.

Chapter 4 provides the outcomes of a multiple-case study in which we were especially interested in projects that targeted flow. Here, we investigated the underlying principles of variability and buffers in the context of Lean improvement projects and show the complexities inherent to these principles. After this third piece of research, we began to observe a pattern. In all of the studies so far, we had been confronted with limitations in explaining our observations based on the trade-off between time and capacity buffers.

In our final research project, reported in Chapter 5, we investigated the buffer trade-off and further developed the concept of quality buffers identified by Hopp *et al.* (2007). We elaborate on their ideas for this additional buffer in service environments, and unravel the underlying buffering mechanisms. The results in Chapter 5 help to explain situations where one fails to observe idle or excess capacity without the presence of obviously waiting patients.

1.3 RESEARCH PROJECT & THESIS OUTLINE

The previous section ended with a brief overview of how the developing thesis focus resulted in four research projects. We now extend this by providing a short introduction to the subsequent chapters.

Chapter 2

In Chapter 2, we focus on performance changes stemming from Lean interventions in a medical laboratory. This research is one of the first to link series of Lean interventions and performance over time. The main research question for this study was: *how do Lean interventions influence time buffer performance over time?* We attempt to answer this research question in a mixed-method longitudinal case study.

The analysis is based on six years of patient-related throughput data, retrieved from a laboratory computer database, to show how throughput time performance evolved over time. In addition, we apply semi-structured interviews to gain a deeper understanding of the Lean interventions. Given the retrospective and longitudinal nature of the study, we use the event history calendar method to enhance the respondents' recall and reliability.

Results provide clarity on the effects of ongoing series of Lean interventions in performance terms in the form of improved throughput times. Additionally, based on the identified Lean interventions that had an impact on throughput time performance, the study offers support for theories emphasizing the importance of reducing variability. However, the outcomes also raise questions regarding the focus of Lean interventions in general: given that only a limited number of these Lean interventions improve throughput performance, what types of issues are addressed?

Chapter 3

In this chapter, we build on the previous research and investigate the focus of Lean interventions initiated by healthcare professionals. Our basic premise is that variability is detrimental to flow performance, and that Lean approaches can be used to reduce variability and thereby improve flow performance. However, it is unclear whether Lean interventions do attempt to reduce variability and consequential buffers, or merely to reduce waste. For this study, we formulated the following research questions: *(1) how should one qualify the focus of existing Lean interventions?; (2) does this focus change, or mature, over time?;* and *(3) does knowledge on buffers and variability influence the intervention focus?* In this research project, we combined a longitudinal field study with an exploratory field quasi-experiment.

In the study, we first classify a large number of interventions from practice, and then carry out an exploratory field quasi-experiment. Our project is one of the few studies that have investigated Lean applications over a lengthy period with a large set of Lean interventions included in the classification. Field experiments in Lean-oriented research are rare, and we see this as a valuable addition to the more commonly used research methods. Through the classification, we establish an overview of the focus of Lean interventions. Semi-structured interviews with practitioners also supported the findings regarding this focus. The exploratory field quasi-experiment was used to investigate whether a lack of knowledge could explain the identified focus.

Results show a limited application of Lean principles, and indicate that expanding knowledge on other, underused, Lean aspects could broaden the focus. This study contributes to the discussion in the literature on the maturity of Lean in the healthcare field. Moreover, it leads to questions concerning the 'straightforwardness' of buffer exchange in relation to variability reduction.

Chapter 4

Prompted by our earlier studies, this research investigates the interaction between buffers and variability in patient flow. To successfully improve flow in healthcare, a good understanding of the relationship linking variability, buffers, and patient flows is essential. Healthcare is considered a service-focused environment with only limited options to buffer variability, namely through either time or capacity buffers. This limitation increases the importance of the variability – buffer relationship.

Here, we highlight complexities related to the variability – buffer relationship in patient flows. To investigate these issues, we formulated two main research questions: *(1) how do buffers and variability interact in a healthcare environment?* and *(2) how does the interchange, or 'trade-off', between buffers manifest itself in practice?* In answering these questions, we carried out a multiple-case study in which we analyze three different hospital departments that had executed flow improvement projects.

The results indicate that the interaction between variability and buffers is less clear-cut than the literature generally assumes. The study highlights several complicating factors that are relevant in flow-oriented Lean projects. As a step in furthering the existing Lean theory on the roles of variability and buffers, our tentative outcomes are translated into a number of propositions. These propositions can be used as the starting point for future research.

Chapter 5

Time and capacity buffer types are commonly identified in service environments. However, during our research, we came across many situations where neither capacity buffers nor waiting time buffers could explain the apparent situation. Based on our earlier studies, this chapter reports on our conceptual study of the roles of quality buffers. In this research, we study the relationship of quality buffers with other buffer types, and consider how such quality buffers would manifest themselves in practice.

The main research questions for this study are: *(1) how should the mechanisms behind quality buffers be typified and explained?* and *(2) how do quality buffers interact with the earlier-identified buffer mechanisms*. We adopt a theoretical approach in formulating answers to these questions, through which we provide a deeper understanding of the mechanisms behind quality buffers.

Our results indicate that the mechanism behind quality buffers involves adjusting processing times. In our study, we explore how these adjustments can be made. The study shows that this mechanism does not necessarily impact on quality, but that this is dependent on the definition of quality adopted. This research provides a more solid theoretical basis for the roles of buffers in Lean applications in healthcare environments and offers a formal definition of quality buffers.

Chapter 6

In this final chapter, we provide a general discussion on the findings presented in this thesis. We provide a short overview of the completed studies, where we highlight the main findings from each study. Next, we discuss our findings in relation to the current literature and highlight its most important implications for both science and practice. Chapter 6 thus provides the link between the chapters. We combine findings from the individual studies and show the relevance of our combined results. Here, we present our adapted Lean maturity model for healthcare environments and discuss the effects of exchanging buffers in healthcare. We finish Chapter 6 with closing comments on our research projects and on the main findings.

A short overview of Dutch healthcare

The Dutch healthcare system is considered one of the best healthcare systems available, and has been frequently ranked first in the Consumer Health Index (Björnberg, 2013). The Dutch healthcare structure provides high quality, safe, and accessible care for the inhabitants of the Netherlands. However, rising demand for care and limited budgets put pressure on the healthcare system and explain the popularity of healthcare improvement initiatives.

Within the Dutch healthcare system, one can distinguish between preventive care, primary care, secondary care, and long-term care. A typical patient's journey starts with a visit to their general practitioner (GP). In those cases where the GP cannot provide the necessary care, a patient is referred to secondary care facilities (often the local hospital). As such, the GP functions as a gatekeeper such that secondary care, such as specialist care in hospitals, is only accessible upon referral by the GP. The gatekeeper role of the GP has a dampening effect on referrals, resulting in a referral rate of four percent (Schäfer *et al.*, 2010). This pre-screening limits the number of people that use specialized care facilities, which helps reduce overall costs. However, even with the dampening effect of GPs, there are still over 25 million hospital referrals a year (2012 data: CBS, 2014), that is 25 million people that require some form of additional diagnosis or treatment.

Obviously, providing high quality treatment to such numbers comes at a high financial cost, and it requires fully dedicated medical personnel.

Maintaining the high level of quality and accessibility of Dutch healthcare is predicted to require increased funding, with estimates reaching levels of 31% of gross domestic product by 2040 (Van der Horst *et al.*, 2011). There are various reasons for this financial pressure, such as the growing elderly population, which is expected to increase healthcare demands in coming years, more treatment possibilities, and better-informed and demanding patients. Furthermore, Dutch healthcare insurers are now allowed to negotiate costs with healthcare providers, making the cost of delivering this high quality care additionally important. Insurers are free to decide not to contract certain providers, for example because of poor standards of delivered care (Daley and Gubb, 2013). Consequently, the healthcare sector is adopting various quality and continuous improvement initiatives to reduce costs.

The interest in improvement initiatives partly helps to explain the popularity of the Lean philosophy. A streamlined process, free of faults and free of unnecessary waiting, is vital. The coming together of these issues underlines the practical relevance of Lean in healthcare. Lean is expected to offer a bottom-up solution to the increasing costs and demands by improving current processes in order to cope with the future demands on healthcare.