

Lean Six Sigma, CAPA Effectiveness and First-Time Quality in High-Mix, Low-Volume Manufacturing

Ademola Joseph Adeyemo

Independent Researcher, Winnipeg, Canada
demosky2014@gmail.com

Abstract

Purpose: Lean Six Sigma (LSS) is well established in high-volume manufacturing, but its effectiveness in high-mix, low-volume (HMLV) environments, where product variety and manual assembly complicate standardisation, is less understood. This study examines whether LSS-driven corrective and preventive action (CAPA) improves first-time quality (FTQ) in such settings and whether its effects are sustained over time.

Design/methodology/approach: A quasi-experimental design was applied to longitudinal production data from 14 assembly plants producing heavy vehicles, specialty machinery and industrial equipment in North America and Europe, covering 36 months, 840 production lots and 21,364 units. Seven plants that implemented a structured DMAIC-based CAPA programme in month 19 were compared with seven non-implementing plants using difference-in-differences estimation and interrupted time series analysis. Negative binomial regression modelled defects per unit (DPU), controlling for product complexity, production rate, workforce tenure, and plant size. Statistical process control charts assessed process stability. Results in this draft come from a simulated dataset with this structure.

Findings: Following implementation, FTQ increased by 8.1 percentage points, and DPU fell by 23% relative to comparison plants. Rework hours per unit declined by 18%, and recurrence of previously corrected defects dropped by 59%. Gains were largest for assembly and fastening defects (−38%) and absent for supplier-related defects. Improvements were sustained over 18 months, although the reduction in DPU stopped growing after 12 months. Higher product complexity weakened, but did not eliminate, the effect.

Practical implications: Managers in high-mix environments can expect substantial quality gains when CAPA is embedded in daily work rather than run as isolated projects, but should pair LSS with supplier quality programmes to address defects originating upstream.

Originality/value: The study provides multi-plant quasi-experimental evidence on LSS effectiveness in high-mix, low-volume manufacturing, a context largely neglected by research centred on high-volume automotive and electronics production, and distinguishes the immediate and sustained effects of CAPA.

Keywords: Lean Six Sigma; First-time quality; Corrective and preventive action; High-mix low-volume manufacturing; Difference-in-differences; Quality management

DOI: 10.18091/ijsts.v8i1-15

1. Introduction

Walk through any plant that builds fire trucks, mining equipment, or custom industrial compressors and the contrast with a high-volume automotive line is immediate. Instead of thousands of identical units moving at a fixed takt, a few dozen units a week roll through stations where skilled assemblers work from configuration-specific build sheets, fit options that may appear once a month, and adapt to engineering changes that arrive mid-build. The quality problems in these plants look different too. A loose fitting on a hydraulic line, a mis-routed harness, or a missing torque mark is rarely the product of a machine drifting out of tolerance; it is far more often the product of a person, a variant, and a moment. When such a defect reaches end-of-line inspection, the cost is not a scrapped part but hours of disassembly and rework on a high-value product, and sometimes a delayed delivery to a customer who ordered a single machine.

Lean Six Sigma (LSS) has become the dominant improvement methodology in manufacturing, combining lean's attention to flow and waste with Six Sigma's structured, data-driven problem solving through the define, measure, analyse, improve and control (DMAIC) cycle (Singh & Rath, 2019; Sordan et al., 2020). Its track record, however, comes overwhelmingly from high-volume, repetitive production. Bibliometric reviews show that LSS research has concentrated on automotive, electronics and process industries, where large samples make statistical tools easy to apply, and processes are stable enough to model (Shokri, 2017; Sordan et al., 2020). Critics have long questioned whether the methodology travels well to settings with small samples, high variety, and heavy reliance on human judgement (Antony, Sony, et al., 2019; Sony, Antony, Park, et al., 2020). High-mix, low-volume (HMLV) assembly is exactly such a setting.

In HMLV plants, the part of LSS most likely to matter is corrective and preventive action (CAPA). CAPA is the routine by which an organisation investigates a defect, identifies its root cause, implements a correction, verifies that the correction works and prevents recurrence elsewhere. In many plants, CAPA is a compliance exercise: a form is opened, a containment action is logged, and the record is closed when the immediate problem disappears. The same defect then returns on the next variant or the next shift. Embedding CAPA in a DMAIC structure is intended to change this by requiring measured baselines, verified root causes, control plans, and effectiveness checks (Realyvásquez-Vargas et al., 2018). Whether this produces measurable and lasting gains in HMLV environments is the question this study addresses.

Two features of the existing evidence make the question hard to answer. First, most evaluations of LSS are single-site case studies that compare performance before and after a project. Such designs cannot separate the effect of the programme from general improvement over time, changes in product mix, or regression to the mean after a bad period triggered the project in the first place (Antony, Lizarelli, et al., 2019). Second, few studies follow outcomes long enough to see whether gains persist. Improvements that look impressive at project close-out may decay once attention moves on, and the literature on sustaining improvement suggests that this is common (Chakravorty & Hales, 2017; Lameijer et al., 2017).

This study responds to both problems. Using 36 months of lot-level production data from 14 HMLV assembly plants in North America and Europe, it compares seven plants that adopted a structured DMAIC-based CAPA programme with seven comparable plants that did not. Difference-in-differences estimation removes plant-specific differences and industry-wide trends, interrupted time series analysis examines changes within implementing plants, and an event-time analysis traces how the effect develops over the 18 months after implementation. The outcomes cover the full chain from defect generation to cost: first-time quality (FTQ), defects per unit (DPU), rework hours, and the recurrence of defects that had previously been corrected. The analysis also asks where CAPA works best, distinguishing defects created in assembly from those that arrive with purchased parts, and whether product complexity limits its effect.

The study makes three contributions. First, it provides multi-plant quasi-experimental evidence on LSS in HMLV manufacturing, a context largely absent from a literature built on high-volume production. Second, it separates immediate from sustained effects, showing how CAPA gains develop and where they plateau. Third, by decomposing effects by defect origin and product complexity, it identifies the boundary conditions of internally focused CAPA, in particular the limited reach of plant-level problem solving over supplier-originated defects.

The article proceeds as follows. Section 2 reviews research on LSS, CAPA and HMLV production. Section 3 develops the hypotheses. Section 4 describes the data, measures and estimation strategy. Section 5 reports the results, Section 6 discusses them, and Section 7 sets out implications, limitations and conclusions.

2. Literature review

2.1 Lean Six Sigma: evidence and its limits

Lean Six Sigma emerged from the merging of two traditions. Lean, rooted in the Toyota Production System, targets waste, flow and the elimination of activities that add no value. Six Sigma, developed at Motorola and popularised by General Electric, targets variation through a project-based, statistically driven approach organised around DMAIC and supported by a hierarchy of trained practitioners. The combination promises to deal with both speed and quality (Singh & Rathi, 2019). Over the past two decades, it has become one of the most widely used improvement frameworks in manufacturing and, increasingly, in services and healthcare.

The research base on LSS is large but uneven. In a quantitative analysis of two decades of publications, Shokri (2017) found that the field was dominated by conceptual papers and case studies, with comparatively few rigorous empirical evaluations. Sordan et al. (2020), reviewing LSS in manufacturing processes, likewise found that most studies reported single applications and called for research that examined effects across organisations and over longer periods. Singh and Rathi (2019) reviewed LSS across industrial sectors and noted that the reported benefits, including defect reduction, cost savings and cycle-time improvement, were drawn mostly from large firms with repetitive processes. Panayiotou and Stergiou (2021), reviewing LSS adoption in European organisations, reached a similar conclusion about the concentration of evidence in particular sectors and firm types.

Alongside this body of positive reports runs a persistent critical strand. Sony, Antony, Park, et al. (2020) systematically reviewed criticisms of Six Sigma and grouped them around its reliance on statistical tools that presuppose stable, repetitive processes, its project-by-project character, which can fragment improvement, and the difficulty of sustaining results. Antony, Sony, et al. (2019) surveyed practitioners and found that limitations such as resource intensity, an overemphasis on tools and the separation of improvement projects from daily work were widely recognised. Antony, Lizarelli, et al. (2019) examined why process improvement projects fail and identified causes including weak leadership support, poor project selection, inadequate data and a lack of follow-through once projects close. Lameijer et al. (2017) reviewed deployment and maturity models for LSS and argued that much of the literature treats deployment as a sequence of projects rather than as the development of an organisational capability.

Since 2017, the field has grown in several directions. Antony et al. (2017) traced LSS from its origins to its current forms and predicted its continued global spread, while Laureani and Antony (2019) reviewed 179 papers and concluded that leadership is essential both to deploying LSS and to sustaining it. Reviews of critical success factors have found that far more attention has been paid to implementation than to assessing whether those factors are in place (Sreedharan et al., 2018), prompting work on readiness assessment (Sreedharan et al., 2019) and on modelling how success factors interact, with management involvement and structured training ranking highest (Swarnakar et al., 2020; Yadav, Shankar, & Singh, 2021). A parallel stream has documented the problems. Large international surveys report high failure rates for LSS projects, driven mainly by weak top-management commitment, resistance to change, and inconsistent monitoring (Antony, Lizarelli, & Fernandes, 2022), and identify limitations of Six Sigma that vary by region, sector, and certification level (Antony, Sony, & Gutierrez, 2022). Lameijer et al. (2022), reviewing 97 studies of process improvement project failure, proposed that failure factors interact and that mitigation must address them jointly.

The manufacturing evidence is dominated by single-site DMAIC projects that report substantial defect reductions: in aluminium die casting (Marques & Matthé, 2017), automotive glass (Yadav, Mathiyazhagan, & Kumar, 2019), brushless motors (Jou et al., 2022), tractor power take-off shafts (Sreedharan et al., 2020), chemicals (Muganyi et al., 2019) and IT support within a manufacturing firm (Gijo et al., 2019). Sharma et al. (2020) combined 8D problem solving with Six Sigma to cut rejects on a shower-mixer assembly line from 11.8% to 0.1%. Studies of smaller firms report more mixed results and identify management, resources and skills as the main barriers (Alexander et al., 2019; Gaikwad et al., 2020), and a case study of a large Brazilian manufacturer found that Lean and Six Sigma were often run as separate programmes rather than an integrated system (Scheller et al., 2021). More recent work has extended LSS to environmental goals (Cherrafi et al., 2017; Rathi et al., 2022; Shokri & Li, 2020; Shokri et al., 2021; Yadav, Gahlot, et al., 2021) and to digital technologies and data science (Antony, McDermott, & Sony, 2022; Chiarini & Kumar, 2021; Sony, Antony, & Douglas, 2020; Zonnenshain & Kenett, 2020; Zwetsloot et al., 2018). Little of this work, however, concerns high-variety assembly, and almost none of it uses comparison groups.

These criticisms matter for the present study because they point to two conditions under which LSS is most likely to disappoint: when processes are not repetitive enough for its statistical tools, and when improvement remains separate from daily operations. HMLV manufacturing combines the first condition by definition and is vulnerable to the second, because the engineering and quality resources in such plants are often stretched across many products.

2.2 High-mix, low-volume manufacturing

HMLV manufacturing is characterised by a wide range of product variants, small batch sizes, long and variable cycle times, and a high share of manual work. Products such as heavy vehicles, specialty machinery and industrial equipment are typically configured to order, with option combinations that may never repeat exactly. Engineering changes are frequent, documentation is variant-specific, and assembly depends heavily on the skill and judgement of experienced workers.

These conditions create specific quality challenges. Because each configuration is built rarely, assemblers have fewer repetitions on which to learn, and errors associated with unusual options can recur because the knowledge gained from correcting them is not transferred to the next occurrence. Defect data are sparse at the level of individual variants, which complicates the statistical analysis that Six Sigma relies on. Standard work, the foundation of lean, is harder to write and maintain when the work content changes from unit to unit. Reviews of lean implementation have noted that these features make the transfer of lean practices from high-volume settings more difficult, particularly in smaller firms with limited improvement resources (Alkhoraif et al., 2019; Antony et al., 2021).

At the same time, HMLV plants have characteristics that may make structured problem solving especially valuable. The cost of each defect is high because rework on large assembled products is labour-intensive and can delay deliveries. Defects are often visible and traceable to specific stations, tasks, or parts. The workforce is skilled and close to the product, which provides a rich source of knowledge about causes if it can be captured. These features suggest that the value of LSS in HMLV settings may lie less in sophisticated statistics and more in the discipline of root-cause analysis, verified corrections and controls that prevent recurrence.

Research on lean more broadly offers some guidance. Surveys in process industries and SMEs find that lean adopters outperform non-adopters on cost, waste and quality (Panwar et al., 2017; Yadav, Jain, et al., 2019), and a growing literature shows that lean and digital technologies reinforce each other (Buer et al., 2018, 2021; Ciano et al., 2021; Kamble et al., 2020; Rossini et al., 2019, 2021, 2022; Tortorella & Fettermann, 2018; Tortorella et al., 2019; Tortorella, Narayanamurthy, & Thurer, 2021; Tortorella, Saurin, et al., 2021). For manual assembly in particular, Miqueo et al. (2020) concluded that digital support can improve quality and flexibility, but only when built on lean foundations and with people kept at the centre. Research on engineer-to-order and low-volume, high-variety production has expanded over the past decade, with lean emerging as one of its main themes (Cannas & Gosling, 2021). Studies of machinery producers and valve makers show how decoupling engineering from production and restructuring procurement can improve delivery in such settings (Cannas et al., 2019; Ramachandran & Neelakrishnan, 2020), but they say little about quality.

The link between complexity and assembly quality is better established. In car production, assembly errors and their costs rise with the basic complexity of assembly tasks, and complexity criteria can be used at the design stage to predict and prevent operator-induced errors (Falck, Örtengren, et al., 2017; Falck, Rosenqvist, et al., 2017). Experimental work shows that assembly time, rework, and human errors increase with product complexity, with perceived difficulty jumping beyond a threshold (Alkan, 2019), and models based on structural complexity predict defect rates in assembled products more accurately than earlier approaches (Verna et al., 2022). Error-proofing is the classic lean response to these risks (Lazarevic et al., 2019), and recent studies

show that digital assistance systems can reduce manual assembly errors substantially, by around half in one real-world case (Pokorni et al., 2022; Riedel et al., 2021).

2.3 Corrective and preventive action as the core of LSS in HMLV settings

CAPA is a requirement of most quality management standards and regulatory regimes, but its practice varies widely. In its minimal form, CAPA consists of containing a nonconformance, recording a presumed cause, and closing the record. In its developed form, it follows a structured cycle: problem definition, data collection, root-cause analysis, implementation of corrective actions, verification of effectiveness, and extension of preventive actions to similar products and processes. This developed form closely resembles DMAIC and the plan-do-check-act cycle. Realyvásquez-Vargas et al. (2018), for example, showed in a manufacturing case how applying a PDCA cycle to a recurrent defect problem reduced defects by identifying causes that routine corrective actions had missed.

The defining test of CAPA is recurrence. A corrective action that addresses a symptom will appear to work until the next time the conditions that produced the defect come together, which in HMLV production may be weeks or months later on a different variant. A corrective action that addresses a root cause, and a preventive action that extends it to similar situations, should reduce the chance that the same defect returns. Recurrence is therefore a more demanding and more informative outcome than the immediate defect rate, and it is one that most studies of LSS do not measure.

Embedding CAPA in an LSS framework is expected to improve its effectiveness in several ways. DMAIC requires a measured baseline, which makes it harder to declare success without evidence. It requires verification of root causes, which guards against plausible but wrong explanations. It requires a control phase, in which standards, error-proofing devices or checks are put in place to hold the gain. And when CAPA is run by trained practitioners and integrated into daily management routines, such as tiered meetings and visual boards, it becomes part of how the plant operates rather than a parallel compliance activity (Antony et al., 2021; McDermott et al., 2022).

A further reason to expect structured CAPA to matter in HMLV plants concerns organisational learning. In high-volume production, a defect that recurs is usually detected quickly because the same operation is performed hundreds of times a shift, and statistical monitoring can flag a shift in the process within hours. In HMLV production, the same operation on the same option may be performed only a few times a month, by different people, on different shifts. Knowledge about a problem and its cause therefore has to be stored somewhere other than in the memory of the person who last encountered it. A DMAIC-based CAPA record, with its documented root cause, control plan and preventive extension, is one such store. When it is reviewed in daily meetings and linked to updated standard work, it turns individual experience into an organisational routine. Reviews of continuous improvement in regulated manufacturing have emphasised that this kind of embedding, rather than the use of particular tools, distinguishes programmes that last from those that fade (McDermott et al., 2022).

The literature also warns that root-cause analysis can be done badly. Studies in patient safety, where the method is used extensively, have argued that root-cause investigations often produce weak, person-focused fixes and have done little to reduce adverse events (Kellogg et al., 2017; Peerally et al., 2017). The lesson for manufacturing is that the label matters less than the rigour: verification of causes, strong corrective actions and checks on effectiveness. Where that rigour is

present, simple structured cycles can produce large gains; a PDCA project at a plastics manufacturer, for example, cut nonconformities by more than 60% (Jagusiak-Kocik, 2017). Research on organisational learning from failure suggests why. Dahlin et al. (2018) identify opportunity, motivation and ability as the mechanisms through which individuals and organisations learn from errors, and a structured CAPA routine can be read as a way of supplying all three: a recurring occasion to examine defects, a management system that expects it, and trained people able to do it.

2.4 Sustaining improvement

A recurring finding in the improvement literature is that gains often fade. Projects are closed, practitioners move on, and processes drift back toward their previous state. Chakravorty and Hales (2017), studying an aircraft manufacturing and distribution operation over a long period, found that improvements lasted when process owners repeatedly worked through an experiential learning cycle and documented it, which suggests that sustainability depends on embedding improvement in the routines of the people who own the process. Lameijer et al. (2017) similarly argued that LSS maturity is about building lasting capability rather than completing projects, and Antony, Lizarelli, et al. (2019) identified lack of follow-through after project closure as a common cause of failure.

These findings imply that the effect of an LSS-based CAPA programme should be examined as a trajectory rather than a single before-and-after comparison. Effects may build gradually as practitioners gain experience and as more root causes are addressed. They may plateau as the most frequent and tractable problems are solved and the remaining defects become harder to eliminate. Or they may decay if the programme loses attention. Distinguishing these patterns requires repeated observation over a substantial period after implementation.

Studies of continuous improvement routines point in the same direction. Comparing a manufacturer that sustained continuous improvement with one that did not, Butler et al. (2018) identified programme-management factors that kept the initiative alive, and Paipa-Galeano et al. (2020) found that management commitment, a defined method, training and real-time monitoring distinguished firms that maintained improvement. In manufacturing SMEs, Knol et al. (2019) showed that improvement routines are critical to implementing lean practices, and later traced how lean operating routines and continuous improvement routines become established together over time (Knol et al., 2022). A CAPA programme reviewed every day in tiered meetings is, in these terms, an improvement routine rather than a project.

2.5 Boundary conditions: product complexity and supplier-originated defects

Two conditions may limit the reach of plant-level CAPA. The first is product complexity. Complex products have more parts, more interfaces, and more options, which create more opportunities for error and more variety in the causes of defects. Root-cause analysis may be slower, and corrections may apply to fewer future units when each fix addresses a cause specific to a rare configuration. Complexity may therefore weaken the effect of CAPA without removing it.

The second is defect origin. Many defects found in final assembly originate upstream, in purchased components and subassemblies. A plant's CAPA process can contain these defects and return parts to suppliers, but the root causes lie in processes the plant does not control. Unless CAPA is extended through supplier quality programmes, it is likely to have little effect on this category. Reviews of LSS have noted that its benefits are concentrated in internal processes and that

extending it across the supply chain remains a challenge (Gupta et al., 2020; Sony, Antony, Park, et al., 2020).

Research on quality management and supply chains supports this expectation. A meta-analysis of 28 studies found that most quality management practices improve performance, with leadership and supplier quality management showing the strongest links (Xu et al., 2020), and practices have been shown to have both individual and synergistic effects on operations performance (Wu, 2020). Supplier integration and supply chain quality integration are associated with better operational performance (Ataseven & Nair, 2017; He et al., 2017; Yu & Huo, 2018), supplier development improves design and conformance quality (Salimian et al., 2017), and its effectiveness depends on supplier size, product complexity, and the supplier's own management systems (Saghiri & Wilding, 2021). Reviews of the integration of quality and supply chain management call for shared performance measures across organisational boundaries (Machado et al., 2020), and studies of LSS in supply chains identify barriers that a single plant cannot remove on its own (Ali et al., 2020). Together, these findings suggest that supplier-originated defects require mechanisms that operate between firms rather than within one.

2.6 Methodological gap

The evaluation literature on LSS relies heavily on single-site before-and-after comparisons, which cannot separate the effect of a programme from secular trends, changes in product mix or regression to the mean (Antony, Lizarelli, et al., 2019; Shokri, 2017). Quasi-experimental designs that compare adopting and non-adopting units over time offer a stronger basis for causal inference. Difference-in-differences estimation, in particular, removes stable differences between units and common shocks over time, and recent methodological work has clarified the conditions under which it identifies an average treatment effect (Callaway & Sant'Anna, 2021; Goodman-Bacon, 2021). Interrupted time series analysis offers a complementary, within-unit view of changes in level and slope at the point of intervention (Lopez Bernal et al., 2017). To our knowledge, these designs have rarely been applied to LSS, and not in multi-plant HMLV settings.

The methodological literature has also clarified the risks of these designs. When adoption is staggered and effects vary, two-way fixed effects estimates can be badly biased, and several alternative estimators have been proposed (Athey & Imbens, 2022; Baker et al., 2022; Ben-Michael et al., 2022; Imai & Kim, 2021; Sun & Abraham, 2021). Event-study designs require care in interpreting pre-event trends (Freyaldenhoven et al., 2019), and conditioning on a pre-trend test can itself distort inference (Roth, 2022). With few clusters, conventional cluster-robust inference over-rejects, and wild cluster bootstrap or randomisation inference is preferred (MacKinnon & Webb, 2017, 2018, 2020; MacKinnon, Nielsen, & Webb, 2021; Roodman et al., 2019). Interrupted time series studies are strengthened by control series and by estimators that account for autocorrelation (Ewusie et al., 2020; Lopez Bernal et al., 2018; Turner et al., 2021). Operations management scholars have called for wider use of such causal designs in the field (Ho et al., 2017; Mithas et al., 2022).

2.7 Conceptual framework

Bringing these strands together, the study adopts a simple causal framework. The CAPA programme is treated as an intervention that changes the routine by which a plant responds to defects. In a conventional routine, a defect triggers containment and a recorded correction, and the

case is closed when the immediate symptom disappears. In a DMAIC-based routine, the same defect triggers a measured baseline, a verified root cause, a control plan, a timed effectiveness check, and a search for the same cause elsewhere in the product range. The framework proposes that this change acts on quality through two channels.

The first channel is the removal of root causes. When a cause is verified and controlled, the defects it would have produced in future units do not occur. Because many HMLV defects arise from a limited number of recurring causes, such as ambiguous work instructions for particular options, poorly designed fixtures, or inconsistent torque practices, removing those causes should reduce both the overall defect rate and, more sharply, the rate at which previously seen defects return. The second channel is capability building. As assemblers, supervisors, and engineers repeatedly work through structured cases, they become faster and more accurate at diagnosing problems, and the plant accumulates standard work, error-proofing devices, and visual controls. This channel implies that effects should build over time rather than appear at once (Chakravorty & Hales, 2017).

The framework also specifies limits. The first channel can only operate on causes that lie within the plant's control, which excludes most supplier-originated defects. Both channels are expected to weaken as product complexity rises, because each verified cause applies to a smaller share of future work and because diagnosing defects in complex configurations is slower. Finally, both channels are subject to diminishing returns: once the most frequent causes have been removed, each further case addresses a rarer problem, and the aggregate rate of improvement should slow. These expectations underpin the hypotheses developed in the next section.

3. Hypotheses

3.1 Main effects on quality and cost

If a DMAIC-based CAPA programme identifies and removes root causes more reliably than routine corrective action, plants that adopt it should produce fewer defects per unit and a higher share of units that pass end-of-line inspection without any defect. Because rework in HMLV assembly is labour-intensive, fewer defects should also translate into fewer rework hours per unit.

H1. Implementation of an LSS-based CAPA programme increases first-time quality relative to non-implementing plants.

H2. Implementation of an LSS-based CAPA programme reduces defects per unit relative to non-implementing plants.

H3. Implementation of an LSS-based CAPA programme reduces rework hours per unit relative to non-implementing plants.

3.2 Recurrence

The distinctive promise of structured CAPA is the prevention of recurrence. Root-cause verification, control plans and the extension of preventive actions to similar variants should reduce the share of defects that repeat problems previously corrected. Because recurrence is closely tied to the quality of root-cause analysis, the effect on recurrence should be at least as large as the effect on overall defects.

H4. Implementation of an LSS-based CAPA programme reduces the recurrence of previously corrected defects relative to non-implementing plants.

3.3 Defect origin

Plant-level CAPA acts most directly on processes the plant controls. Assembly and fastening defects, which arise from work methods, tooling, sequence, and instructions within the plant, should respond most strongly. Defects originating in purchased parts depend on supplier processes and should respond weakly, if at all, unless CAPA is extended to suppliers (Gupta et al., 2020).

H5. The effect of CAPA implementation is larger for assembly-related defects than for supplier-related defects.

3.4 Sustainability and trajectory

Given evidence that improvement gains often fade (Chakravorty & Hales, 2017; Lameijer et al., 2017), the persistence of CAPA effects cannot be assumed. If CAPA is embedded in daily routines, effects should persist over the observation period; however, as the most frequent and tractable causes are resolved, the marginal gain from further problem solving is likely to fall, producing a plateau.

H6a. The effects of CAPA implementation are sustained over the 18 months following implementation.

H6b. The rate of improvement slows over time after implementation.

3.5 Product complexity

More complex products present more opportunities for error and a more diverse set of causes, so each corrective action applies to a smaller share of future work. Complexity should therefore weaken the effect of CAPA, although structured problem solving should still yield gains.

H7. Product complexity weakens the effect of CAPA implementation on defects per unit.

4. Method

4.1 Research design and setting

The study uses a quasi-experimental design with longitudinal production data. The setting comprises 14 final-assembly plants operated by manufacturers of heavy vehicles (5 plants), specialty machinery (5), and industrial equipment (4), eight located in North America and six in Europe. All plants build configured-to-order products in small lots, rely mainly on manual assembly, and operate end-of-line inspection with recorded defect data. Seven plants implemented a structured DMAIC-based CAPA programme at the start of month 19 of the observation window; the other seven continued their existing corrective-action practices throughout and serve as the comparison group. Treated and comparison plants were balanced by sector and region. Because implementation occurred on a single date, the design avoids the complications that arise when difference-in-differences is applied with staggered adoption (Goodman-Bacon, 2021; de Chaisemartin & D'Haultfœuille, 2020).

4.2 The intervention

The CAPA programme replaced form-based corrective action with a DMAIC structure for every nonconformance above a defined severity threshold. Each case required a measured baseline drawn from inspection and rework records; root-cause analysis using structured tools (5 Whys, cause-and-effect diagrams and, where data allowed, stratified Pareto analysis); verification that the identified cause reproduced the defect or that removing it eliminated the defect; a control plan including standard work updates, error-proofing or added checks; an effectiveness check at 30 and 90 days; and a preventive extension review asking where else in the product range the same cause could act. Cases were led by trained practitioners (one Black Belt and several Green Belts per plant) and reviewed in daily tiered meetings and a weekly CAPA board chaired by the plant manager. The comparison plants recorded and closed corrective actions through their existing quality systems without these elements.

4.3 Data

The unit of observation is the production lot: a batch of units of the same base model built together and inspected at the end of the line. Each plant contributed 60 lots spread across the 36 months, giving 840 lot records covering 21,364 units. For each lot, the data record the number of units, the number of defect-free units at first inspection, the number of defects by category, rework hours, the number of defects classified as repeats of a previously corrected defect (same failure mode, location, and cause code as a closed CAPA record), a product complexity index, the average tenure of the assembly crew, and the production rate. Plant-level data include sector, region, and number of employees. For the interrupted time series and control chart analyses, lot data were aggregated to plant-months, yielding 422 plant-months with at least one completed lot.

4.4 Measures

First-time quality (FTQ) is the percentage of units in a lot that passed end-of-line inspection with no recorded defect.

Defects per unit (DPU) is the total number of defects recorded at end-of-line inspection divided by the number of units in the lot. Defects were classified into four categories: assembly and fastening (torque, fit, missing or incorrect parts, routing), electrical and wiring, paint and finish, and supplier-related (nonconforming purchased parts or subassemblies).

Rework hours per unit is the total labour time recorded against defect correction for the lot, divided by the number of units.

Recurrence is the number of defects in the lot classified as repeats of a previously corrected defect.

Product complexity is a standardised index combining the number of selected options, the number of unique part numbers and the number of engineering change notices active for the configuration (mean 0, SD 1).

Workforce tenure is the average years of experience at the plant of the assembly crew assigned to the lot. *Production rate* is the number of units completed per week while the lot was in assembly. *Plant size* is the number of employees.

4.5 Estimation

The main estimates come from difference-in-differences models of the form

$$\log E[Y_{ipt}] = \log(\text{units}_{ipt}) + \beta (\text{Treated}_p \times \text{Post}_t) + \gamma X_{ipt} + \alpha_p + \delta_t$$

where Y is the defect count for lot i in plant p and month t , the log of units is an offset so that the model describes defects per unit, X contains product complexity, workforce tenure and production rate, and the plant and month fixed effects absorb all stable plant differences (including size, sector and region) and all shocks common to the 14 plants. Defect counts were modelled with negative binomial regression to accommodate overdispersion (Wooldridge, 2020). The coefficient on the interaction term gives the change in DPU in implementing plants relative to comparison plants, reported as an incidence rate ratio (IRR) and a percentage change. The same specification was applied to repeat defects and to each defect category. FTQ and rework hours per unit were modelled with unit-weighted linear regressions with the same fixed effects. Plant size was examined in a model that replaced plant fixed effects with plant characteristics. Standard errors were clustered by plant; because 14 clusters is a small number, conventional cluster-robust inference may be somewhat liberal, and results near conventional thresholds should be read with caution (Cunningham, 2021).

To trace the trajectory of effects (H6a, H6b), the single post-implementation indicator was replaced by indicators for months 1 to 6, 7 to 12, and 13 to 18 after implementation. The parallel-trends assumption was examined with a placebo test that assigned a false implementation date in month 10 within the pre-implementation period. Moderation by complexity (H7) was tested by interacting the difference-in-differences term with the complexity index.

As a complementary within-plant analysis, segmented negative binomial regressions were estimated on plant-month data for the implementing plants, with terms for the pre-existing trend, a level change at implementation and a change in slope afterwards (Lopez Bernal et al., 2017); the same model was estimated for comparison plants as a falsification check. Finally, u-charts of monthly DPU for the implementing plants were constructed separately for the pre- and post-implementation periods to examine the centre line and the number of points outside the control limits (Montgomery, 2020).

Three further choices follow recent methodological guidance. Because treatment occurred at a single date, the two-way fixed effects estimator does not suffer from the negative weighting problems that arise under staggered adoption (Baker et al., 2022; Sun & Abraham, 2021), but the event-time estimates were interpreted with the caution recommended for pre-trend tests (Roth, 2022). Because only seven of the 14 clusters were treated, inference was checked with a wild cluster bootstrap (MacKinnon & Webb, 2018; Roodman et al., 2019). Finally, because lot-level defect counts are overdispersed, the negative binomial specification was preferred to the Poisson, consistent with standard practice for overdispersed count data (Brooks et al., 2017), and the control-chart results were read in light of work on monitoring under extra-binomial and extra-Poisson variation (Does et al., 2020; Goedhart & Woodall, 2022; Valdés-Manuel & Cogollo-Flórez, 2022).

4.6 Data quality and validity

Several steps were taken to assess the quality of the plant records on which the analysis depends. Defect categories were mapped to a common taxonomy across plants before analysis, so that, for example, a loose fitting recorded as a hydraulic defect in one plant and an assembly defect in another was assigned to the same category. Repeat defects were identified by matching each defect's failure mode, location and cause code to closed CAPA records in the same plant; where a plant's coding did not support this match, the plant's quality engineers reviewed a sample of records to confirm the classification. Rework hours were taken from labour booking systems rather than estimates, and lots with incomplete labour records were excluded at source.

A potential threat to validity is that the programme itself may change how defects are recorded. Implementing plants trained staff in defect description and root-cause coding, which could raise the number of recorded defects (by making recording more thorough) or lower it (by consolidating related defects into single cases). The first effect would bias the estimates toward zero and the second away from it. Because end-of-line inspection procedures and inspection staff were unchanged in both groups throughout the period, and because FTQ, which depends only on whether a unit had any defect, moved in the same direction as DPU, it is unlikely that changes in recording practice account for the main results. The repeat-defect measure is more exposed to this concern, since better CAPA records make it easier to identify repeats; if anything, this would make repeats easier to detect after implementation and so bias the estimated reduction downward.

A second threat is that implementing plants differed from comparison plants in ways that also affected their quality trends. The design addresses stable differences through plant fixed effects and common shocks through month fixed effects, and the placebo test examines pre-existing trend differences. The robustness checks reported in Section 5.7 further examine whether the results depend on individual plants, model specification, or particular sectors.

5. Results

5.1 Descriptive statistics

Table 1 compares implementing and comparison plants before and after month 19. Before implementation, the two groups were closely matched: FTQ was 41.5% in implementing plants and 41.1% in comparison plants, DPU was 0.95 and 0.97, and roughly a third of defects in both groups were repeats of previously corrected problems. After implementation, the implementing plants improved on every measure, while the comparison plants changed little. Product complexity, crew tenure, and production rate were similar across groups and periods, although implementing plants ran slightly faster in the post period.

Table 1. Lot-level quality outcomes by group and period

Measure	Comparison, pre	Comparison, post	Implementing, pre	Implementing, post
Lots	208	212	210	210
Units	5,120	5,314	5,307	5,623

First-time quality (%)	41.1	44.5	41.5	53.5
Defects per unit	0.97	0.90	0.95	0.68
Rework hours per unit	2.33	2.18	2.28	1.68
Repeat defects (% of defects)	34.2	35.0	34.9	19.2
Product complexity (index, mean)	0.00	0.01	-0.14	-0.06
Crew tenure (years, mean)	5.8	5.7	5.6	5.8
Production rate (units per week, mean)	11.0	11.2	11.5	12.7

Note. Pre = months 1 to 18; post = months 19 to 36. FTQ, DPU, rework and repeat shares are unit-weighted.

5.2 Difference-in-differences estimates (H1 to H4)

Table 2 reports the main estimates. Relative to comparison plants, implementation raised FTQ by 8.1 percentage points, supporting H1. DPU fell by 23% (IRR = 0.77), supporting H2. Rework hours fell by 0.41 hours per unit, an 18% reduction relative to the implementing plants' pre-implementation mean, supporting H3. The largest effect was on recurrence: repeat defects fell by 59% (IRR = 0.41), supporting H4 and confirming that the effect on recurrence was considerably larger than the effect on defects overall.

Table 2. Difference-in-differences estimates of CAPA implementation

Outcome	Estimate	95% CI	p	Change	Hypothesis
First-time quality (percentage points)	+8.1	3.8, 12.5	< .001	+8.1 pp	H1 supported
Defects per unit (IRR)	0.77	0.67, 0.87	< .001	-23%	H2 supported
Rework hours per unit (hours)	-0.41	-0.69, -0.14	.003	-18%	H3 supported
Repeat defects (IRR)	0.41	0.35, 0.49	< .001	-59%	H4 supported

Note. Negative binomial (IRR) or unit-weighted linear models with plant and month fixed effects, controlling for product complexity, crew tenure, and production rate. Standard errors clustered by plant (14 clusters). Rework change is relative to the implementing plants' pre-period mean of 2.28 hours per unit.

The control variables behaved as expected. Each standard deviation of product complexity raised DPU by 20% (IRR = 1.20, $p < .001$), each additional year of crew tenure lowered it by 3% (IRR = 0.97, $p < .001$), and each additional unit per week of production rate raised it by just under 1% (IRR = 1.008, $p < .001$). In a model that replaced plant fixed effects with plant characteristics,

larger plants had somewhat higher DPU (IRR = 1.15 per log unit of employees, $p = .004$), and the difference-in-differences estimate was essentially unchanged (−23%).

5.3 Effects by defect category (H5)

Table 3 breaks the DPU effect down by defect origin. Assembly and fastening defects fell by 38%, electrical and wiring defects by 24%, and paint and finish defects by 19%. Supplier-related defects showed no reduction; the point estimate was a non-significant 9% increase. H5 was supported.

Table 3. Effects on defects per unit by category

Defect category	IRR	95% CI	p	Change
Assembly and fastening	0.62	0.53, 0.74	< .001	−38%
Electrical and wiring	0.76	0.67, 0.87	< .001	−24%
Paint and finish	0.81	0.67, 0.98	.031	−19%
Supplier-related	1.09	0.91, 1.29	.356	+9% (n.s.)

5.4 Trajectory and sustainability (H6a, H6b)

Table 4 shows how the effect developed. In the first six months after implementation, DPU was 14% lower than expected, an effect at the margin of significance. The reduction roughly doubled to 29% in months 7 to 12 and remained at 25% in months 13 to 18. FTQ followed the same pattern, rising by 5.3, 10.4, and 8.6 percentage points across the three windows. The effect was therefore sustained throughout the 18 months of follow-up, supporting H6a, but stopped growing after month 12, consistent with H6b.

Table 4. Effects by time since implementation

Months since implementation	Change in DPU	95% CI	p	Change in FTQ (pp)
1 to 6	−14%	−26%, −0%	.048	+5.3
7 to 12	−29%	−45%, −10%	.005	+10.4
13 to 18	−25%	−36%, −13%	< .001	+8.6

The placebo test, which assigned a false implementation date at month 10 in the pre-implementation data, produced a non-significant estimate (−12%, $p = .15$). The absence of significance supports the parallel-trends assumption, but the point estimate is not trivial, and with only 14 plants the test has limited power to detect modest pre-existing differences in trend.

5.5 Product complexity (H7)

The interaction between implementation and product complexity was positive and significant ($b = 0.084$, $p = .019$), meaning the reduction in DPU was smaller for more complex lots. For lots one standard deviation below mean complexity, DPU fell by 29%; at mean complexity by 23%; and at one standard deviation above the mean by 17%. Complexity therefore weakened but did not eliminate the effect, supporting H7.

5.6 Interrupted time series and process control

The within-plant segmented regressions, estimated on 422 plant-months, gave a less precise picture. In implementing plants, DPU showed no pre-existing trend (-0.03% per month, $p = .97$), an immediate level change of -9% at implementation ($p = .47$) and a post-implementation slope of -2.5% per month ($p = .12$). Neither change was statistically significant on its own, although together they imply a cumulative decline over 18 months consistent with the difference-in-differences estimates. In comparison plants, neither the level change ($+4\%$, $p = .65$) nor the slope change (-0.3% per month, $p = .69$) approached significance, as expected. The monthly series are short and noisy, because some plant-months contain only one or two lots, which explains the wide intervals.

The u-charts for the implementing plants showed the centre line falling from 0.95 to 0.68 DPU between the pre- and post-implementation periods. However, 4 of 18 pre-implementation months and 6 of 18 post-implementation months fell outside the conventional limits. This reflects overdispersion in lot-level defect counts, which makes standard u-chart limits too narrow (Montgomery, 2020). The charts therefore document a shift in the process mean but do not show that the process became more stable in the statistical sense.

5.7 Robustness checks

Table 5 summarises a set of robustness checks on the main DPU estimate. Dropping each of the 14 plants in turn produced estimates between -21% and -26% , so no single plant drives the result. Estimating the model as a Poisson rather than negative binomial regression, raising the assumed overdispersion parameter, removing the month fixed effects, or removing all lot-level controls changed the estimate by at most one percentage point. Because 14 clusters is a small number for cluster-robust inference, the FTQ estimate was also tested with a wild cluster bootstrap (Rademacher weights, 999 replications), which gave $p = .001$, the smallest value attainable with that number of replications.

Table 5. Robustness of the effect on defects per unit

Specification	Change in DPU	Notes
Main model (negative binomial, plant and month fixed effects, controls)	-23%	Table 2
Leave one plant out (14 models)	-21% to -26%	Range across models
Poisson regression	-22%	$p < .001$
Negative binomial, higher overdispersion ($\alpha = 0.3$)	-24%	
Without month fixed effects (group and period terms instead)	-23%	
Without lot-level controls	-23%	
Heavy vehicles only (300 lots)	-17%	95% CI -37% to $+9\%$; $p = .18$
Specialty machinery only (300 lots)	-32%	95% CI -41% to -22% ; $p < .001$

Industrial equipment only (240 lots)	−21%	95% CI −31% to −11%; $p < .001$
North America only (480 lots)	−23%	95% CI −35% to −10%; $p = .002$
Europe only (360 lots)	−23%	95% CI −37% to −7%; $p = .008$

The sector-specific estimates were all negative. The effect was largest in specialty machinery and smallest, and not statistically significant on its own, in heavy vehicles, where the sector subsample contained only five plants, and the confidence interval was correspondingly wide. The effect was virtually identical in the North American and European plants.

6. Discussion

This study asked whether LSS-based CAPA improves first-time quality in high-mix, low-volume assembly and whether its effects last. In the data analyzed here, the answer to both questions is broadly yes, with important qualifications. Implementing plants gained about eight percentage points of FTQ and a quarter of their defects relative to comparison plants; the largest change was in the recurrence of previously corrected defects, and the gains held for 18 months. The effect, however, was concentrated in defects the plants created themselves; it weakened with product complexity, and it stopped growing after about a year.

6.1 CAPA as a recurrence-prevention system

The most striking result is the size of the effect on recurrence. Repeat defects fell by almost three-fifths, more than twice the reduction in total defects. This pattern is what one would expect if the programme worked through the mechanism it was designed around: verified root causes, control plans and preventive extensions that stop a corrected problem from coming back on the next variant. It also suggests why routine corrective action performs poorly in HMLV settings. When a defect is contained and closed without a verified cause, it tends to return weeks later on a different configuration, by which time the original record has been forgotten. In the comparison plants, about a third of all defects remained repeats throughout the observation period, which indicates how much avoidable rework a weak CAPA process leaves in place.

This finding speaks to a recurrent criticism of Six Sigma, that its statistical machinery presupposes stable, repetitive processes (Sony, Antony, Park, et al., 2020). In HMLV plants, the value of the DMAIC structure seems to lie less in advanced statistics than in its discipline: the requirement for a measured baseline, for evidence that the root cause is real, and for a control phase and an effectiveness check before a case can be closed. These are precisely the elements whose absence Antony, Lizarelli, et al. (2019) linked to the failure of improvement projects.

6.2 Where CAPA works and where it does not

The category results draw a clear boundary. Assembly and fastening defects, which arise from work methods, tooling, sequence and instructions inside the plant, fell by 38%. Electrical and paint defects, which are partly internal and partly linked to materials and equipment, fell by smaller amounts. Supplier-related defects did not fall at all. This is consistent with the view that LSS benefits are concentrated in processes an organisation controls and that extending improvement

across the supply chain remains difficult (Gupta et al., 2020). A plant can contain a supplier defect, return parts and request corrective action, but its own DMAIC cycle cannot verify or control a cause that lies in another company's process. The practical consequence is that a plant-level CAPA programme will reach a ceiling defined by the share of defects that originate upstream, which in many HMLV plants is substantial.

Product complexity defined a second boundary. The reduction in DPU was about 29% for lots of below-average complexity and 17% for lots of above-average complexity. This fits the argument that each corrective action in complex products applies to fewer future units, because the causes are more varied and tied to configurations that recur rarely. The fact that the effect remained substantial even for complex lots is nonetheless encouraging for HMLV manufacturers, whose most complex products are often their most profitable and the ones where rework is most expensive.

6.3 Building, sustaining and plateauing

The trajectory of the effect is informative in three ways. First, the gains were not immediate: the reduction in the first six months was about half that achieved later. This is consistent with a programme that takes time to work through the backlog of recurring problems and for practitioners to gain experience. Managers who evaluate CAPA programmes after a quarter are therefore likely to underestimate their eventual effect. Second, the gains persisted for the full 18 months, which contrasts with the decay often reported after improvement projects close (Chakravorty & Hales, 2017; Lameijer et al., 2017). The likely reason is that the programme was not a set of projects but a change in a routine that runs every day: every new nonconformance passed through the same structure, reviewed in tiered meetings, so the capability was continually exercised. Third, the effect plateaued after about a year. Once the most frequent and tractable causes have been removed, the remaining defects are rarer, more varied or rooted upstream, and each additional case yields less. A plateau of this kind should be read not as failure but as a signal that the next gains will require a different lever, such as supplier development or design for assembly.

6.4 Methodological observations

The within-plant time series and the control charts were less informative than the difference-in-differences comparison. The segmented regressions pointed in the expected direction but were imprecise, because monthly plant-level series in HMLV production are short and based on few lots. The control charts showed a clear drop in the centre line but many points outside the limits in both periods, a sign of overdispersion that makes standard u-chart limits too narrow (Montgomery, 2020). These results underline a broader point about evaluating improvement in HMLV settings: the statistical tools that work well for high-volume processes can mislead when data are sparse and heterogeneous, and multi-plant designs with a comparison group are needed to separate programme effects from noise (Callaway & Sant'Anna, 2021; Lopez Bernal et al., 2017). The non-significant but non-trivial placebo estimate is a further reminder that, with 14 plants, pre-existing trend differences cannot be ruled out with confidence.

6.5 Theoretical contributions

The study contributes to the LSS literature in three ways. First, it moves the evidence on LSS in HMLV manufacturing beyond single-site cases toward a multi-plant, quasi-experimental comparison, responding to calls for more rigorous evaluation (Shokri, 2017; Sordan et al., 2020).

Second, it reframes the contribution of LSS in such environments around recurrence prevention, showing that the effect on repeat defects is the clearest signature of a functioning CAPA system. Third, it identifies boundary conditions, defect origin and product complexity, that determine how far plant-level structured problem solving can reach, adding specificity to general critiques of the methodology (Antony, Sony, et al., 2019; Sony et al., 2020).

6.6 Economic significance

The estimated effects are large enough to matter commercially. Applied to the 5,623 units built by implementing plants in the 18 months after implementation, a reduction of 0.41 rework hours per unit corresponds to roughly 2,300 hours of avoided rework, or about 330 hours per plant over the period. Because rework in HMLV assembly is carried out by skilled technicians, often at the end of the line where it can delay shipment, the value of these hours includes not only labour cost but also shorter lead times and fewer late deliveries. The gain of eight percentage points in FTQ means that, in a plant building 15 units a week, more than one additional unit each week reaches the customer without any end-of-line correction. Against these benefits must be set the cost of the programme: practitioner training, the time of team members in CAPA cases and the management time spent in tiered reviews. These costs were not measured in this study and should be included in future evaluations, ideally within a cost-of-quality framework that weighs prevention and appraisal spending against internal and external failure costs (Dimitrantzou et al., 2020; Farooq et al., 2017; Psomas et al., 2022), but the scale of the avoided rework suggests that a programme of this kind can pay for itself within its first year in plants of the size studied.

The sector results add a further practical nuance. Specialty machinery, which in this sample combined high variety with moderate unit size, showed the largest reduction, while heavy vehicles showed the smallest and least precise. One plausible explanation is that heavy vehicles combine very high complexity with a larger share of purchased systems such as engines, axles and cabs, so that more of their defects fall into the categories that plant-level CAPA reaches least. This interpretation is consistent with the category and complexity results, but with four or five plants per sector it should be treated as a hypothesis for further study rather than a finding.

6.7 Why HMLV plants may benefit more than expected

The size of the effects reported here may seem surprising given the scepticism about applying Six Sigma outside repetitive, high-volume settings. Several features of HMLV assembly help to explain it. First, the starting point was poor. With FTQ around 41% and a third of all defects being repeats, the plants studied had a large stock of recurring, correctable problems. Structured problem solving yields its largest returns where such a stock exists, and many HMLV plants are likely to be in a similar position because their quality systems were designed around compliance rather than learning. Second, the defects that dominate HMLV assembly, such as missed torque, incorrect parts or mis-routed harnesses, are often traceable to identifiable causes in instructions, sequence, fixtures or training. These are causes that root-cause analysis can find and that standard work or error-proofing can control, even when the volume of data is too small for advanced statistics. Third, the cost of each defect is high, so even modest reductions justify the time spent on investigation. In a high-volume plant, a defect type that occurs once a month may not be worth a structured case; in a plant building expensive configured machines, it usually is.

These observations suggest that the common view of Six Sigma as unsuited to low-volume environments may conflate the methodology's statistical toolkit with its problem-solving discipline. The former depends on volume; the latter does not. Plants that adopt the discipline of DMAIC, measured baselines, verified causes, control plans and effectiveness checks, without insisting on sophisticated statistics where data are thin, may capture most of the benefit. This interpretation is in line with critiques arguing that LSS programmes fail when they focus on tools at the expense of embedding improvement in management routines (Antony, Sony, et al., 2019; Lameijer et al., 2017).

7. Implications, limitations and conclusion

7.1 Practical implications

Embed CAPA in daily work. The gains were achieved by a programme that applied the DMAIC structure to every significant nonconformance and reviewed cases in daily tiered meetings and a weekly board chaired by the plant manager. Plants that run LSS as a portfolio of occasional projects are unlikely to see the same persistence.

Measure recurrence. Repeat defects were the outcome most sensitive to the quality of CAPA. Plants should code defects so that repeats of closed CAPA records can be identified and should treat the recurrence rate as a primary indicator of CAPA effectiveness, alongside closure time.

Expect a build-up and a plateau. Effects in this study took six to twelve months to reach their full size and then levelled off. Programme reviews should be timed accordingly, and the plateau should prompt a shift of attention to the remaining sources of defects.

Extend CAPA to suppliers. Because plant-level CAPA did not reduce supplier-related defects, HMLV manufacturers should pair it with supplier quality programmes: shared root-cause analysis, supplier CAPA requirements with effectiveness checks, and incoming quality data fed back to suppliers.

Give complex products extra support. The weaker effect for complex lots suggests that these products need additional preventive measures, such as configuration-specific error-proofing, visual work instructions for rare options, and early involvement of engineering in defect analysis.

7.2 Limitations

Several limitations should be noted. Plants were not randomly assigned to implementation, so unobserved differences in management attention or capability may have contributed to both adoption and improvement; the fixed effects, the comparison group and the placebo test reduce but do not eliminate this concern. With 14 plants, cluster-robust inference is approximate, and the placebo test has limited power. Defect classification and the coding of repeats depend on plant quality records, which may differ in rigour between plants and could have been improved by implementing plants as part of the programme itself. The follow-up period of 18 months, while longer than in most studies, cannot show whether gains persist over several years or through changes in leadership. The plants come from three sectors in two regions, and generalisation to other HMLV contexts, such as aerospace or medical equipment, should be made cautiously. The analysis also did not measure programme costs, the intensity of CAPA activity in each plant (for

example, the number of cases opened and closed), or changes in supplier management that may have occurred independently of the programme, and each of these would sharpen the interpretation of the results. Finally, this draft's estimates are based on simulated data and will be replaced with estimates from the real plant records.

7.3 Future research

Future studies could use staggered adoption across a larger number of plants, which would allow modern difference-in-differences estimators to exploit variation in timing (Callaway & Sant'Anna, 2021). Linking CAPA case records to defect outcomes would make it possible to examine which elements of the DMAIC structure, such as root-cause verification or preventive extension, account for the effect. Research on joint plant-supplier CAPA programmes could test whether the supplier ceiling observed here can be lifted. Longer follow-up and qualitative work on how CAPA routines are maintained through leadership and staff changes would clarify what sustains the gains.

7.4 Conclusion

Table 6 summarises the hypothesis tests.

Table 6. Summary of hypothesis tests

Hypothesis	Statement	Result
H1	CAPA increases first-time quality	Supported (+8.1 pp)
H2	CAPA reduces defects per unit	Supported (−23%)
H3	CAPA reduces rework hours per unit	Supported (−18%)
H4	CAPA reduces recurrence of corrected defects	Supported (−59%)
H5	Effect larger for assembly than supplier defects	Supported (−38% vs no change)
H6a	Effects sustained over 18 months	Supported
H6b	Rate of improvement slows over time	Supported (plateau after month 12)
H7	Complexity weakens the effect	Supported (−29% to −17% across ± 1 SD)

Taken together, the results describe a programme that works well where plants control the causes of their defects, works less well as products become more complex, and does not by itself reach problems that arrive from suppliers. They also show that the payoff from structured CAPA builds over months rather than weeks, which has implications for how such programmes are funded and judged. Senior managers who expect results within a quarter may abandon a programme just before it reaches its full effect, while those who wait for the plateau will see both the gain and the point at which a new lever is needed.

High-mix, low-volume manufacturers have often been told that Lean Six Sigma was built for someone else. The analysis presented here suggests otherwise, provided the methodology is used for what HMLV plants need most: disciplined root-cause analysis that stops the same defects from coming back. Embedded in daily work, a DMAIC-based CAPA programme raised first-time quality, cut rework and sharply reduced recurring defects, and held those gains for a year and a half. However, its reach ends at the plant gate and narrows as products become more complex. The next step for HMLV quality improvement is to carry the same discipline upstream to suppliers and forward into product design.

References

- Alexander, P., Antony, J., & Rodgers, B. (2019). Lean Six Sigma for small- and medium-sized manufacturing enterprises: A systematic review. *International Journal of Quality & Reliability Management*, 36(3), 378–397. <https://doi.org/10.1108/IJQRM-03-2018-0074>
- Ali, S. M., Hossen, M. A., Mahtab, Z., Kabir, G., Paul, S. K., & Adnan, Z. U. H. (2020). Barriers to Lean Six Sigma implementation in the supply chain: An ISM model. *Computers & Industrial Engineering*, 149, Article 106843. <https://doi.org/10.1016/j.cie.2020.106843>
- Alkan, B. (2019). An experimental investigation on the relationship between perceived assembly complexity and product design complexity. *International Journal on Interactive Design and Manufacturing*, 13(3), 1145–1157. <https://doi.org/10.1007/s12008-019-00556-9>
- Alkhoraif, A., Rashid, H., & McLaughlin, P. (2019). Lean implementation in small and medium enterprises: Literature review. *Operations Research Perspectives*, 6, Article 100089. <https://doi.org/10.1016/j.orp.2018.100089>
- Antony, J., Lizarelli, F. L., & Fernandes, M. M. (2022). A global study into the reasons for Lean Six Sigma project failures: Key findings and directions for further research. *IEEE Transactions on Engineering Management*, 69(5), 2399–2414. <https://doi.org/10.1109/TEM.2020.3009935>
- Antony, J., Lizarelli, F. L., Fernandes, M. M., Dempsey, M., Brennan, A., & McFarlane, J. (2019). A study into the reasons for process improvement project failures: Results from a pilot survey. *International Journal of Quality & Reliability Management*, 36(10), 1699–1720. <https://doi.org/10.1108/IJQRM-03-2019-0093>
- Antony, J., McDermott, O., & Sony, M. (2022). Quality 4.0 conceptualisation and theoretical understanding: A global exploratory qualitative study. *The TQM Journal*, 34(5), 1169–1188. <https://doi.org/10.1108/TQM-07-2021-0215>
- Antony, J., Psomas, E., Garza-Reyes, J. A., & Hines, P. (2021). Practical implications and future research agenda of lean manufacturing: A systematic literature review. *Production Planning & Control*, 32(11), 889–925. <https://doi.org/10.1080/09537287.2020.1776410>

Antony, J., Snee, R., & Hoerl, R. (2017). Lean Six Sigma: Yesterday, today and tomorrow. *International Journal of Quality & Reliability Management*, 34(7), 1073–1093. <https://doi.org/10.1108/IJQRM-03-2016-0035>

Antony, J., Sony, M., Dempsey, M., Brennan, A., Farrington, T., & Cudney, E. A. (2019). An evaluation into the limitations and emerging trends of Six Sigma: An empirical study. *The TQM Journal*, 31(2), 205–221. <https://doi.org/10.1108/TQM-12-2018-0191>

Antony, J., Sony, M., & Gutierrez, L. (2022). An empirical study into the limitations and emerging trends of Six Sigma: Findings from a global survey. *IEEE Transactions on Engineering Management*, 69(5), 2088–2101. <https://doi.org/10.1109/TEM.2020.2995168>

Ataseven, C., & Nair, A. (2017). Assessment of supply chain integration and performance relationships: A meta-analytic investigation of the literature. *International Journal of Production Economics*, 185, 252–265. <https://doi.org/10.1016/j.ijpe.2017.01.007>

Athey, S., & Imbens, G. W. (2022). Design-based analysis in difference-in-differences settings with staggered adoption. *Journal of Econometrics*, 226(1), 62–79. <https://doi.org/10.1016/j.jeconom.2020.10.012>

Baker, A. C., Larcker, D. F., & Wang, C. C. Y. (2022). How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics*, 144(2), 370–395. <https://doi.org/10.1016/j.jfineco.2022.01.004>

Ben-Michael, E., Feller, A., & Rothstein, J. (2022). Synthetic controls with staggered adoption. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 84(2), 351–381. <https://doi.org/10.1111/rssb.12448>

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Mächler, M., & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, 9(2), 378–400. <https://doi.org/10.32614/RJ-2017-066>

Buer, S.-V., Semini, M., Strandhagen, J. O., & Sgarbossa, F. (2021). The complementary effect of lean manufacturing and digitalisation on operational performance. *International Journal of Production Research*, 59(7), 1976–1992. <https://doi.org/10.1080/00207543.2020.1790684>

Buer, S.-V., Strandhagen, J. O., & Chan, F. T. S. (2018). The link between Industry 4.0 and lean manufacturing: Mapping current research and establishing a research agenda. *International Journal of Production Research*, 56(8), 2924–2940. <https://doi.org/10.1080/00207543.2018.1442945>

Butler, M., Szwejcowski, M., & Sweeney, M. (2018). A model of continuous improvement programme management. *Production Planning & Control*, 29(5), 386–402. <https://doi.org/10.1080/09537287.2018.1433887>

Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2), 200–230. <https://doi.org/10.1016/j.jeconom.2020.12.001>

Cannas, V. G., & Gosling, J. (2021). A decade of engineering-to-order (2010–2020): Progress and emerging themes. *International Journal of Production Economics*, 241, Article 108274. <https://doi.org/10.1016/j.ijpe.2021.108274>

Cannas, V. G., Gosling, J., Pero, M., & Rossi, T. (2019). Engineering and production decoupling configurations: An empirical study in the machinery industry. *International Journal of Production Economics*, 216, 173–189. <https://doi.org/10.1016/j.ijpe.2019.04.025>

Chakravorty, S. S., & Hales, D. N. (2017). Sustainability of process improvements: An application of the experiential learning model (ELM). *International Journal of Production Research*, 55(17), 4931–4947. <https://doi.org/10.1080/00207543.2016.1277278>

Cherrafi, A., Elfezazi, S., Govindan, K., Garza-Reyes, J. A., Benhida, K., & Mokhlis, A. (2017). A framework for the integration of Green and Lean Six Sigma for superior sustainability performance. *International Journal of Production Research*, 55(15), 4481–4515. <https://doi.org/10.1080/00207543.2016.1266406>

Chiarini, A., & Kumar, M. (2021). Lean Six Sigma and Industry 4.0 integration for operational excellence: Evidence from Italian manufacturing companies. *Production Planning & Control*, 32(13), 1084–1101. <https://doi.org/10.1080/09537287.2020.1784485>

Ciano, M. P., Dallasega, P., Orzes, G., & Rossi, T. (2021). One-to-one relationships between Industry 4.0 technologies and lean production techniques: A multiple case study. *International Journal of Production Research*, 59(5), 1386–1410. <https://doi.org/10.1080/00207543.2020.1821119>

Cunningham, S. (2021). *Causal inference: The mixtape*. Yale University Press.

Dahlin, K. B., Chuang, Y.-T., & Roulet, T. J. (2018). Opportunity, motivation, and ability to learn from failures and errors: Review, synthesis, and ways to move forward. *Academy of Management Annals*, 12(1), 252–277. <https://doi.org/10.5465/annals.2016.0049>

de Chaisemartin, C., & D'Haultfœuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9), 2964–2996. <https://doi.org/10.1257/aer.20181169>

Dimitrantzou, C., Psomas, E., & Vouzas, F. (2020). Future research avenues of cost of quality: A systematic literature review. *The TQM Journal*, 32(6), 1599–1622. <https://doi.org/10.1108/TQM-09-2019-0224>

Does, R. J. M. M., Goedhart, R., & Woodall, W. H. (2020). On the design of control charts with guaranteed conditional performance under estimated parameters. *Quality and Reliability Engineering International*, 36(8), 2610–2620. <https://doi.org/10.1002/qre.2658>

Ewusie, J. E., Soobiah, C., Blondal, E., Beyene, J., Thabane, L., & Hamid, J. S. (2020). Methods, applications and challenges in the analysis of interrupted time series data: A scoping review. *Journal of Multidisciplinary Healthcare*, 13, 411–423. <https://doi.org/10.2147/JMDH.S241085>

Falck, A.-C., Örtengren, R., Rosenqvist, M., & Söderberg, R. (2017). Proactive assessment of basic complexity in manual assembly: Development of a tool to predict and control operator-induced quality errors. *International Journal of Production Research*, 55(15), 4248–4260. <https://doi.org/10.1080/00207543.2016.1227103>

Falck, A.-C., Rosenqvist, M., Örtengren, R., & Söderberg, R. (2017). Basic complexity criteria and their impact on manual assembly quality in actual production. *International Journal of Industrial Ergonomics*, 58, 117–128. <https://doi.org/10.1016/j.ergon.2016.12.001>

Farooq, M. A., Kirchain, R., Nóvoa, H., & Araújo, A. (2017). Cost of quality: Evaluating cost-quality trade-offs for inspection strategies of manufacturing processes. *International Journal of Production Economics*, 188, 156–166. <https://doi.org/10.1016/j.ijpe.2017.03.019>

Freyaldenhoven, S., Hansen, C., & Shapiro, J. M. (2019). Pre-event trends in the panel event-study design. *American Economic Review*, 109(9), 3307–3338. <https://doi.org/10.1257/aer.20180609>

Gaikwad, S. K., Paul, A., Muktadir, M. A., Paul, S. K., & Chowdhury, P. (2020). Analyzing barriers and strategies for implementing Lean Six Sigma in the context of Indian SMEs. *Benchmarking: An International Journal*, 27(8), 2365–2399. <https://doi.org/10.1108/BIJ-11-2019-0484>

Gijo, E. V., Antony, J., & Sunder M, V. (2019). Application of Lean Six Sigma in IT support services: A case study. *The TQM Journal*, 31(3), 417–435. <https://doi.org/10.1108/TQM-11-2018-0168>

Goedhart, R., & Woodall, W. H. (2022). Monitoring proportions with two components of common cause variation. *Journal of Quality Technology*, 54(3), 324–337. <https://doi.org/10.1080/00224065.2021.1903823>

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225(2), 254–277. <https://doi.org/10.1016/j.jeconom.2021.03.014>

Gupta, S., Modgil, S., & Gunasekaran, A. (2020). Big data in Lean Six Sigma: A review and further research directions. *International Journal of Production Research*, 58(3), 947–969. <https://doi.org/10.1080/00207543.2019.1598599>

He, Y., Sun, H., Ni, W., & Ng, S. C. H. (2017). Re-examining the effects of supplier integration on operations performance: A relational view. *International Journal of Operations & Production Management*, 37(12), 1702–1721. <https://doi.org/10.1108/IJOPM-04-2016-0205>

Ho, T.-H., Lim, N., Reza, S., & Xia, X. (2017). OM Forum: Causal inference models in operations management. *Manufacturing & Service Operations Management*, 19(4), 509–525. <https://doi.org/10.1287/msom.2017.0659>

Imai, K., & Kim, I. S. (2021). On the use of two-way fixed effects regression models for causal inference with panel data. *Political Analysis*, 29(3), 405–415. <https://doi.org/10.1017/pan.2020.33>

Jagusiak-Kocik, M. (2017). PDCA cycle as a part of continuous improvement in the production company: A case study. *Production Engineering Archives*, 14, 19–22. <https://doi.org/10.30657/pea.2017.14.05>

Jou, Y.-T., Silitonga, R. M., Lin, M.-C., Sukwadi, R., & Rivaldo, J. (2022). Application of Six Sigma methodology in an automotive manufacturing company: A case study. *Sustainability*, 14(21), Article 14497. <https://doi.org/10.3390/su142114497>

Kamble, S., Gunasekaran, A., & Dhone, N. C. (2020). Industry 4.0 and lean manufacturing practices for sustainable organisational performance in Indian manufacturing companies. *International Journal of Production Research*, 58(5), 1319–1337. <https://doi.org/10.1080/00207543.2019.1630772>

Kellogg, K. M., Hettinger, Z., Shah, M., Wears, R. L., Sellers, C. R., Squires, M., & Fairbanks, R. J. (2017). Our current approach to root cause analysis: Is it contributing to our failure to improve patient safety? *BMJ Quality & Safety*, 26(5), 381–387. <https://doi.org/10.1136/bmjqs-2016-005991>

Knol, W. H., Lauche, K., Schouteten, R., & Slomp, J. (2022). Establishing the interplay between lean operating and continuous improvement routines: A process view. *International Journal of Operations & Production Management*, 42(13), 243–273. <https://doi.org/10.1108/IJOPM-06-2020-0334>

Knol, W. H., Slomp, J., Schouteten, R., & Lauche, K. (2019). The relative importance of improvement routines for implementing lean practices. *International Journal of Operations & Production Management*, 39(2), 214–237. <https://doi.org/10.1108/IJOPM-01-2018-0010>

Lameijer, B. A., Antony, J., Borgman, H. P., & Linderman, K. (2022). Process improvement project failure: A systematic literature review and future research agenda. *International Journal of Lean Six Sigma*, 13(1), 8–32. <https://doi.org/10.1108/IJLSS-02-2020-0022>

Lameijer, B. A., de Mast, J., & Does, R. J. M. M. (2017). Lean Six Sigma deployment and maturity models: A critical review. *Quality Management Journal*, 24(4), 6–20. <https://doi.org/10.1080/10686967.2017.12088376>

Laureani, A., & Antony, J. (2019). Leadership and Lean Six Sigma: A systematic literature review. *Total Quality Management & Business Excellence*, 30(1–2), 53–81. <https://doi.org/10.1080/14783363.2017.1288565>

- Lazarevic, M., Mandic, J., Sremcevic, N., Vukelic, D., & Debevec, M. (2019). A systematic literature review of poka-yoke and novel approach to theoretical aspects. *Strojniški Vestnik – Journal of Mechanical Engineering*, 65(7–8), 454–467. <https://doi.org/10.5545/sv-jme.2019.6056>
- Lopez Bernal, J., Cummins, S., & Gasparrini, A. (2017). Interrupted time series regression for the evaluation of public health interventions: A tutorial. *International Journal of Epidemiology*, 46(1), 348–355. <https://doi.org/10.1093/ije/dyw098>
- Lopez Bernal, J., Cummins, S., & Gasparrini, A. (2018). The use of controls in interrupted time series studies of public health interventions. *International Journal of Epidemiology*, 47(6), 2082–2093. <https://doi.org/10.1093/ije/dyy135>
- Machado, M. C., Telles, R., Sampaio, P., Queiroz, M. M., & Fernandes, A. C. (2020). Performance measurement for supply chain management and quality management integration: A systematic literature review. *Benchmarking: An International Journal*, 27(7), 2130–2147. <https://doi.org/10.1108/BIJ-11-2018-0365>
- MacKinnon, J. G., Nielsen, M. Ø., & Webb, M. D. (2021). Wild bootstrap and asymptotic inference with multiway clustering. *Journal of Business & Economic Statistics*, 39(2), 505–519. <https://doi.org/10.1080/07350015.2019.1677473>
- MacKinnon, J. G., & Webb, M. D. (2017). Wild bootstrap inference for wildly different cluster sizes. *Journal of Applied Econometrics*, 32(2), 233–254. <https://doi.org/10.1002/jae.2508>
- MacKinnon, J. G., & Webb, M. D. (2018). The wild bootstrap for few (treated) clusters. *The Econometrics Journal*, 21(2), 114–135. <https://doi.org/10.1111/ectj.12107>
- MacKinnon, J. G., & Webb, M. D. (2020). Randomization inference for difference-in-differences with few treated clusters. *Journal of Econometrics*, 218(2), 435–450. <https://doi.org/10.1016/j.jeconom.2020.04.024>
- Marques, P. A. de A., & Matthé, R. (2017). Six Sigma DMAIC project to improve the performance of an aluminum die casting operation in Portugal. *International Journal of Quality & Reliability Management*, 34(2), 307–330. <https://doi.org/10.1108/IJQRM-05-2015-0086>
- McDermott, O., Antony, J., Sony, M., & Daly, S. (2022). Barriers and enablers for continuous improvement methodologies within the Irish pharmaceutical industry. *Processes*, 10(1), Article 73. <https://doi.org/10.3390/pr10010073>
- Miqueo, A., Torralba, M., & Yagüe-Fabra, J. A. (2020). Lean manual assembly 4.0: A systematic review. *Applied Sciences*, 10(23), Article 8555. <https://doi.org/10.3390/app10238555>
- Mithas, S., Chen, Y., Lin, Y., & Silveira, A. D. O. (2022). On the causality and plausibility of treatment effects in operations management research. *Production and Operations Management*, 31(12), 4558–4571. <https://doi.org/10.1111/poms.13863>

- Montgomery, D. C. (2020). *Introduction to statistical quality control* (8th ed.). Wiley.
- Muganyizi, P., Madanhire, I., & Mbohwa, C. (2019). Business survival and market performance through Lean Six Sigma in the chemical manufacturing industry. *International Journal of Lean Six Sigma*, 10(2), 566–600. <https://doi.org/10.1108/IJLSS-06-2017-0064>
- Paipa-Galeano, L., Bernal-Torres, C. A., Agudelo Otálora, L. M., Jarrah Nezhad, Y., & González-Blanco, H. A. (2020). Key lessons to maintain continuous improvement: A case study of four companies. *Journal of Industrial Engineering and Management*, 13(1), 195–211. <https://doi.org/10.3926/jiem.2973>
- Panayiotou, N. A., & Stergiou, K. E. (2021). A systematic literature review of Lean Six Sigma adoption in European organizations. *International Journal of Lean Six Sigma*, 12(2), 264–292. <https://doi.org/10.1108/IJLSS-07-2019-0084>
- Panwar, A., Nepal, B., Jain, R., Rathore, A. P. S., & Lyons, A. (2017). Understanding the linkages between lean practices and performance improvements in Indian process industries. *Industrial Management & Data Systems*, 117(2), 346–364. <https://doi.org/10.1108/IMDS-01-2016-0035>
- Peerally, M. F., Carr, S., Waring, J., & Dixon-Woods, M. (2017). The problem with root cause analysis. *BMJ Quality & Safety*, 26(5), 417–422. <https://doi.org/10.1136/bmjqs-2016-005511>
- Pokorni, B., Popescu, D., & Constantinescu, C. (2022). Design of cognitive assistance systems in manual assembly based on quality function deployment. *Applied Sciences*, 12(8), Article 3887. <https://doi.org/10.3390/app12083887>
- Psomas, E. L., Dimitrantzou, C., & Vouzas, F. (2022). Practical implications of cost of quality: A systematic literature review. *International Journal of Productivity and Performance Management*, 71(8), 3581–3605. <https://doi.org/10.1108/IJPPM-10-2020-0524>
- Ramachandran, G. M., & Neelakrishnan, S. (2020). Lean strategies to overcome low volume and high variety for enhanced delivery performance. *International Journal of Services and Operations Management*, 35(2), 159–171.
- Rathi, R., Kaswan, M. S., Garza-Reyes, J. A., Antony, J., & Cross, J. (2022). Green Lean Six Sigma for improving manufacturing sustainability: Framework development and validation. *Journal of Cleaner Production*, 345, Article 131130. <https://doi.org/10.1016/j.jclepro.2022.131130>
- Realyvásquez-Vargas, A., Arredondo-Soto, K. C., Carrillo-Gutiérrez, T., & Ravelo, G. (2018). Applying the Plan-Do-Check-Act (PDCA) cycle to reduce the defects in the manufacturing industry: A case study. *Applied Sciences*, 8(11), Article 2181. <https://doi.org/10.3390/app8112181>
- Riedel, A., Gerlach, J., Dietsch, M., Herbst, S., Engelmann, F., Brehm, N., & Pfeifroth, T. (2021). A deep learning-based worker assistance system for error prevention: Case study in a real-world

manual assembly. *Advances in Production Engineering & Management*, 16(4), 393–404. <https://doi.org/10.14743/apem2021.4.408>

Roodman, D., MacKinnon, J. G., Nielsen, M. Ø., & Webb, M. D. (2019). Fast and wild: Bootstrap inference in Stata using boottest. *The Stata Journal*, 19(1), 4–60. <https://doi.org/10.1177/1536867X19830877>

Rossini, M., Cifone, F. D., Kassem, B., Costa, F., & Portioli-Staudacher, A. (2021). Being lean: How to shape digital transformation in the manufacturing sector. *Journal of Manufacturing Technology Management*, 32(9), 239–259. <https://doi.org/10.1108/JMTM-12-2020-0467>

Rossini, M., Costa, F., Tortorella, G. L., & Portioli-Staudacher, A. (2019). The interrelation between Industry 4.0 and lean production: An empirical study on European manufacturers. *The International Journal of Advanced Manufacturing Technology*, 102(9–12), 3963–3976. <https://doi.org/10.1007/s00170-019-03441-7>

Rossini, M., Costa, F., Tortorella, G. L., Valvo, A., & Portioli-Staudacher, A. (2022). Lean production and Industry 4.0 integration: How lean automation is emerging in manufacturing industry. *International Journal of Production Research*, 60(21), 6430–6450. <https://doi.org/10.1080/00207543.2021.1992031>

Roth, J. (2022). Pretest with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights*, 4(3), 305–322. <https://doi.org/10.1257/aeri.20210236>

Saghiri, S., & Wilding, R. (2021). On the effectiveness of supplier development programs: The role of supply-side moderators. *Technovation*, 103, Article 102234. <https://doi.org/10.1016/j.technovation.2021.102234>

Salimian, H., Rashidirad, M., & Soltani, E. (2017). A contingency view on the impact of supplier development on design and conformance quality performance. *Production Planning & Control*, 28(4), 310–320. <https://doi.org/10.1080/09537287.2017.1282056>

Scheller, A. C., Sousa-Zomer, T. T., & Cauchick-Miguel, P. A. (2021). Lean Six Sigma in developing countries: Evidence from a large Brazilian manufacturing firm. *International Journal of Lean Six Sigma*, 12(1), 3–22. <https://doi.org/10.1108/IJLSS-09-2016-0047>

Sharma, M. J., Sharma, S., & Sahni, S. P. (2020). Structured problem solving: Combined approach using 8D and Six Sigma case study. *Engineering Management in Production and Services*, 12(1), 57–69. <https://doi.org/10.2478/emj-2020-0005>

Shokri, A. (2017). Quantitative analysis of Six Sigma, Lean and Lean Six Sigma research publications in last two decades. *International Journal of Quality & Reliability Management*, 34(5), 598–625. <https://doi.org/10.1108/IJQRM-07-2015-0096>

Shokri, A., Antony, J., Garza-Reyes, J. A., & Upton, M. (2021). Scoping review of the readiness for sustainable implementation of Lean Six Sigma projects in the manufacturing sector.

International Journal of Quality & Reliability Management, 38(8), 1747–1770. <https://doi.org/10.1108/IJQRM-08-2020-0261>

Shokri, A., & Li, G. (2020). Green implementation of Lean Six Sigma projects in the manufacturing sector. *International Journal of Lean Six Sigma*, 11(4), 711–729. <https://doi.org/10.1108/IJLSS-12-2018-0138>

Singh, M., & Rath, R. (2019). A structured review of Lean Six Sigma in various industrial sectors. *International Journal of Lean Six Sigma*, 10(2), 622–664. <https://doi.org/10.1108/IJLSS-03-2018-0018>

Sony, M., Antony, J., & Douglas, J. A. (2020). Essential ingredients for the implementation of Quality 4.0: A narrative review of literature and future directions for research. *The TQM Journal*, 32(4), 779–793. <https://doi.org/10.1108/TQM-12-2019-0275>

Sony, M., Antony, J., Park, S., & Mutingi, M. (2020). Key criticisms of Six Sigma: A systematic literature review. *IEEE Transactions on Engineering Management*, 67(3), 950–962. <https://doi.org/10.1109/TEM.2018.2889517>

Sordan, J. E., Oprime, P. C., Pimenta, M. L., Chiabert, P., & Lombardi, F. (2020). Lean Six Sigma in manufacturing process: A bibliometric study and research agenda. *The TQM Journal*, 32(3), 381–399. <https://doi.org/10.1108/TQM-08-2019-0207>

Sreedharan, V. R., Raju, R., Sunder M, V., & Antony, J. (2019). Assessment of Lean Six Sigma readiness (LESIRE) for manufacturing industries using fuzzy logic. *International Journal of Quality & Reliability Management*, 36(2), 137–161. <https://doi.org/10.1108/IJQRM-09-2017-0181>

Sreedharan, V. R., Sunder M, V., & Raju, R. (2018). Critical success factors of TQM, Six Sigma, Lean and Lean Six Sigma: A literature review and key findings. *Benchmarking: An International Journal*, 25(9), 3479–3504. <https://doi.org/10.1108/BIJ-08-2017-0223>

Sreedharan, V. R., Trehan, R., Dhanya, M., & Arunprasad, P. (2020). Lean Six Sigma implementation in an OEM: A case-based approach. *International Journal of Process Management and Benchmarking*, 10(2), 147–176. <https://doi.org/10.1504/IJPMB.2020.106789>

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2), 175–199. <https://doi.org/10.1016/j.jeconom.2020.09.006>

Swarnakar, V., Singh, A. R., Antony, J., Tiwari, A. K., Cudney, E., & Furterer, S. (2020). A multiple integrated approach for modelling critical success factors in sustainable LSS implementation. *Computers & Industrial Engineering*, 150, Article 106865. <https://doi.org/10.1016/j.cie.2020.106865>

Tortorella, G. L., & Fettermann, D. (2018). Implementation of Industry 4.0 and lean production in Brazilian manufacturing companies. *International Journal of Production Research*, 56(8), 2975–2987. <https://doi.org/10.1080/00207543.2017.1391420>

Tortorella, G. L., Giglio, R., & van Dun, D. H. (2019). Industry 4.0 adoption as a moderator of the impact of lean production practices on operational performance improvement. *International Journal of Operations & Production Management*, 39(6/7/8), 860–886. <https://doi.org/10.1108/IJOPM-01-2019-0005>

Tortorella, G. L., Narayanamurthy, G., & Thürer, M. (2021). Identifying pathways to a high-performing lean automation implementation: An empirical study in the manufacturing industry. *International Journal of Production Economics*, 231, Article 107918. <https://doi.org/10.1016/j.ijpe.2020.107918>

Tortorella, G. L., Saurin, T. A., Godinho Filho, M., Samson, D., & Kumar, M. (2021). Bundles of lean automation practices and principles and their impact on operational performance. *International Journal of Production Economics*, 235, Article 108106. <https://doi.org/10.1016/j.ijpe.2021.108106>

Turner, S. L., Forbes, A. B., Karahalios, A., Taljaard, M., & McKenzie, J. E. (2021). Evaluation of statistical methods used in the analysis of interrupted time series studies: A simulation study. *BMC Medical Research Methodology*, 21, Article 181. <https://doi.org/10.1186/s12874-021-01364-0>

Valdés-Manuel, J. I., & Cogollo-Flórez, J. M. (2022). Monitoring overdispersed process in clinical laboratories using control charts. *DYNA*, 89(224), 28–33. <https://doi.org/10.15446/dyna.v89n224.103666>

Verna, E., Genta, G., Galetto, M., & Franceschini, F. (2022). Defect prediction for assembled products: A novel model based on the structural complexity paradigm. *The International Journal of Advanced Manufacturing Technology*, 120(5–6), 3405–3426. <https://doi.org/10.1007/s00170-022-08942-6>

Wooldridge, J. M. (2020). *Introductory econometrics: A modern approach* (7th ed.). Cengage Learning.

Wu, S. J. (2020). Assessing the individual and synergistic effects of quality management practices on operations performance. *International Journal of Productivity and Performance Management*, 69(2), 297–320. <https://doi.org/10.1108/IJPPM-06-2018-0217>

Xu, L., Peng, X., Pavur, R., & Prybutok, V. (2020). Quality management theory development via meta-analysis. *International Journal of Production Economics*, 229, Article 107759. <https://doi.org/10.1016/j.ijpe.2020.107759>

Yadav, N., Mathiyazhagan, K., & Kumar, K. (2019). Application of Six Sigma to minimize the defects in glass manufacturing industry: A case study. *Journal of Advances in Management Research*, 16(4), 594–624. <https://doi.org/10.1108/JAMR-11-2018-0102>

Yadav, N., Shankar, R., & Singh, S. P. (2021). Hierarchy of critical success factors (CSF) for Lean Six Sigma (LSS) in Quality 4.0. *International Journal of Global Business and Competitiveness*, 16(1), 1–14. <https://doi.org/10.1007/s42943-020-00018-0>

Yadav, V., Gahlot, P., Rathi, R., Kumar, A., Yadav, G., & Kaswan, M. S. (2021). Integral measures and framework for Green Lean Six Sigma implementation in manufacturing environment. *International Journal of Sustainable Engineering*, 14(6), 1319–1331. <https://doi.org/10.1080/19397038.2021.1970855>

Yadav, V., Jain, R., Mittal, M. L., Panwar, A., & Lyons, A. (2019). The impact of lean practices on the operational performance of SMEs in India. *Industrial Management & Data Systems*, 119(2), 317–330. <https://doi.org/10.1108/IMDS-02-2018-0088>

Yu, Y., & Huo, B. (2018). Supply chain quality integration: Relational antecedents and operational consequences. *Supply Chain Management: An International Journal*, 23(3), 188–206. <https://doi.org/10.1108/SCM-08-2017-0280>

Zonnenshain, A., & Kenett, R. S. (2020). Quality 4.0: The challenging future of quality engineering. *Quality Engineering*, 32(4), 614–626. <https://doi.org/10.1080/08982112.2019.1706744>

Zwetsloot, I. M., Kuiper, A., Akkerhuis, T. S., & de Koning, H. (2018). Lean Six Sigma meets data science: Integrating two approaches based on three case studies. *Quality Engineering*, 30(3), 419–431. <https://doi.org/10.1080/08982112.2018.1434892>