

Designing the accident and emergency system: lessons from manufacturing

P Walley

Emerg Med J 2003;20:126–130

Correspondence to:
Mr P Walley, Warwick
Business School, University
of Warwick, CV4 8AA,
UK;
ompw@wbs.warwick.ac.uk

Accepted for publication
10 December 2002

Objectives: To review the literature on manufacturing process design and demonstrate applicability in health care.

Methods: Literature review and application of theory using two years activity data from two healthcare communities and extensive observation of activities over a six week period by seven researchers.

Results: It was possible to identify patient flows that could be used to design treatment processes around the needs of the patient. Some queues are built into existing treatment processes and can be removed by better process design. Capacity imbalance, not capacity shortage, causes some unnecessary waiting in accident and emergency departments.

Conclusions: Clinicians would find that modern manufacturing theories produce more acceptable designs of systems. In particular, good quality is seen as a necessary pre-requisite of fast, efficient services.

It has been widely regarded by many business commentators that the service sector in general has lagged behind in the adoption of management methodologies compared with manufacturing industry.¹ Arguably, health care is one of the least industrialised areas of economic activity. The lack of competitive pressure and resistance from healthcare professionals who baulk against the “McDonaldisation” of healthcare are just two of the many reasons why change has not happened. The resistance is understandable given the poor reputation of modern production methods, including the phenomenon of “karoshi”² or death by overwork! However, the use of externally derived design methods does not automatically lead to the types of repetitive factory process that many clinicians may fear. Process design theory has developed considerably in the past two decades and it is now possible to achieve a design that is closer to meeting the needs of all stakeholders.

This paper reviews the core theories of manufacturing process design and the transfer of those theories into the service sector. The use of relevant manufacturing theory in health care is also identified. The second part of the paper shows how these theories were applied during the preliminary “Warwick report”³ for the Modernisation Agency’s IDEA project, developing the “Ideal Design of Emergency Access”.

CONVENTIONAL MANUFACTURING DESIGN THEORY

It is generally accepted that many manufacturing and service process designs are the consequence of emergent design processes, instead of formal “process engineering” methodologies (see, for example, Hammer⁴). However, operations management texts consistently recommend similar models of how processes should be designed.⁵ Two basic questions are asked before manufacturing operations can be designed appropriately:

1 What performance do I need from my process?

Very few manufacturing processes are designed exclusively to achieve low unit cost. Other aspects of performance often have a greater priority, or have a greater influence on process design. Designers look at factors such as:

- Probable fluctuations in demand

- Seasonal changes in the mix of products
- The need to provide a fast speed of response
- A need to heavily customise the product or offer a “premium” product
- The consequences of delivering defective items
- The impact of delivering the product too late (or too early)

Top manufacturing companies now take their performance to unprecedented levels. Where quality is a key market requirement, product defect rates can be so low that they are measured in parts per million, rather than percentages. Some customers now give suppliers a 15 minute “window” to deliver products, rather than allow them an entire day’s latitude or longer. Flexible systems cope with millions of product combinations or sudden 10-fold increases in demand. These levels of performance have to be designed in, as they cannot be achieved by better supervision of a poor process.

2 What levels of volume and variety are we likely to see?

In conventional manufacturing design theory, the natural levels of variety and the volume of the output make a considerable difference to the high level configuration of resources. Manufacturing processes are often categorised into five main design archetypes, each one corresponding to a particular combination of volume and variety.⁶ Figure 1 shows how process choice relates to volume and variety.

Each process choice represents a different configuration of resources, influencing the layout and flow of work. Two process choices are especially relevant here. The “batch” process, as the name suggests, groups identical products together and keeps every item at one stage in a process until they have all been worked on. A typical batch product spends 90% of its time waiting for the work around it to be processed, before it is moved to the next production stage. The flow of resource within batch processes is chaotic, entailing long travel distances for parts and intermittent, stop-start flow of resources.

Mass processes are designed around the sequence of the production process. The stereotypical mass process will have a linear production line, with smaller travel distances and smoother flow. Conventional mass processes are now seen to

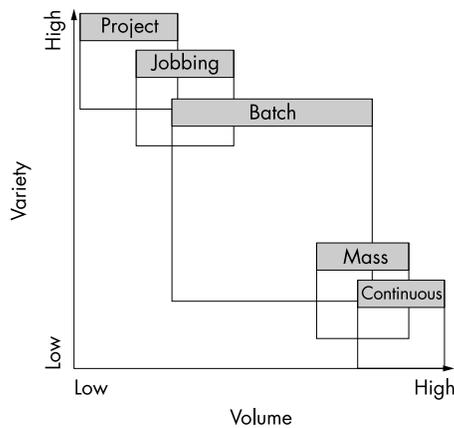


Figure 1 Manufacturing process choice related to volume and variety.

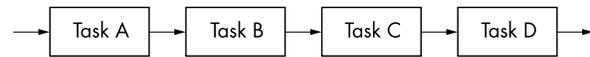
have a number of undesirable characteristics such as uninteresting, repetitive work for employees and a lack of flexibility. Many managers have been reluctant to make the switch from batch to mass production.⁶ However, if processes do not evolve from batch to mass production as their output volume increases, the performance of the system can be seriously affected. In particular, high volume batch processes have considerable problems associated with the long time that work takes to reach the end of the process. The product becomes stuck in excessive and unnecessary queues.

Modern designs of high volume processes have simultaneously improved job design for employees and also improved performance. “Just in time” principles^{7,8} have been used to control the flow through a process. “Demand-pull” mechanisms are used, which limit the amount of work that is waiting by forcing all stages to slow down to the speed of the slowest parts of the system. A key feature is the successful use of employee empowerment, teamworking, and other supportive personnel practices.⁹ Quality is given top priority as it is recognised that high levels of quality must be achieved before a process can become resource efficient. The practices are often called “lean production”,¹⁰⁻¹² which some commentators regard as a paradigm shift in how processes should be designed.

Most manufacturing companies have to produce a variety of products that are in a state of constant change. Products can become obsolete very quickly, with old products regularly replaced by new ones. Manufacturers have to produce all of these products without compromising performance in any one market. One solution has been to divide the product range into smaller groups and produce these in “cellular” operations.^{13,14} It is now believed that 80% of manufacturing organisations in the US use cellular manufacturing.¹⁵ In effect, cellular operations are small scale factories, concentrating on narrower ranges of product, where the process can be designed specifically for a smaller product range. They retain some flexibility, but can achieve some of the efficiency of lean production. Cellular operations are ideal for the implementation of teamwork as they naturally create manageable teams.

The design of cellular operations is a delicate balance between the performance gains that are achieved by better process design and the loss of efficiency caused by ring fenced resources.¹⁶ If managed properly, cellular processes do not improve the performance for one group of products at the expense of others. The system moves towards the best achievable performance for each segment of work. Staff can have more varied jobs, making the work more interesting and less repetitive. Capacity can be matched more easily to demand, as cells can be added incrementally. Lastly, resource balancing is more easily achieved as there tend to be fewer process stages to balance.

A A typical mass process with narrow tasks (one person per task)



B A cellular process with similar capacity, but enlarged tasks

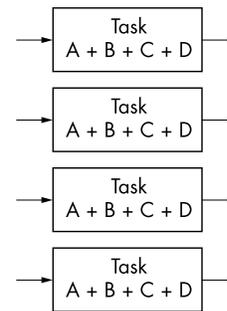


Figure 2 A comparison of conventional and cellular mass process designs.

The concept of manufacturing focus is also relevant to operations design. Skinner¹⁷ argued that manufacturing processes are often given too wide a brief and suffer when trying to achieve irreconcilable and contradictory sets of performance targets. The concept of “plant within plant” was introduced, where smaller scale factories concentrate on single sets of objectives. Focus is often misunderstood to mean high volume, mass production. Some focused factories deliberately operate at low volumes, while others concentrate on particular performance objectives.

Manufacturing principles have been applied to the service sector for a considerable time.^{18,19} It is now recognised that the same problems of poor flexibility and low employee morale that beset manufacturing can be seen in the service sector. McDonald’s inflexible, standardised production system is no longer the exemplar it once was.^{9,20} Quality, flexibility, and fast responsiveness are now seen as better characteristics to embed in a service process.²¹ Consequently, health care is seen as an appropriate application for modern design principles.

APPLICATIONS IN HEALTH CARE

It can be argued that the UK’s healthcare system has exhibited the same reluctance as manufacturing companies to adapt their process choice. As the scale of health services has expanded, the design has not radically changed and is still organised to treat comparatively small numbers of each patient type. This is an issue about the configuration of resources, not the quantity of resources. What once was a process designed to adapt for each almost unique patient now seems to be a poorly standardised system for the higher numbers of similar patients requiring treatment. It is very easy to view the typical NHS hospital as the service equivalent of a batch manufacturing system.²² Most UK hospitals move patients from one department to another in complex, long distance, stop-start flow patterns. In such a system, the scheduling and progressing of patients is extremely difficult. Capacity bottlenecks move, making high utilisation of resources near impossible. It is also easy to view hospitals as unfocused, with the same irreconcilable sets of contradictory performance objectives and organisation priorities. A&E departments clearly have different (and more difficult) targets for speed and flexibility, when compared with many forms of elective treatment.

The concepts of lean production and focus have already been applied in health care. The most widely known application is seen at the Canadian hernia repair centre, Shouldice hospital.^{23,24} The hospital treats non-smoking patients, who are

not overweight, for their inguinal hernias. Historically they have claimed a recurrence rate one tenth that of conventional methods, at a fraction of the normal cost. The design is ruthlessly based around the needs of one type of patient. Patients are encouraged to be ambulatory—they walk to the operating theatre and are expected to climb off the operating table afterwards. The self service restaurant is upstairs, with steps half the normal height to accommodate patients who have been treated just two hours previously. Focus has also been successfully applied to US “surgicentres”.²⁵ There are lean thinking examples in US primary care.²⁶ Just in time is being used for healthcare material management.²⁷

RESEARCH METHODOLOGY

Research was conducted over two months (July to August) 2001 in the two IDEA lead site health regions of North Cheshire and Lewisham. Seven full time researchers were used to gather the data on the two sites. Direct observation of the A&E processes and process changes, together with detailed process mapping, took place at this time. Activity information could be extracted from the IT systems or paper records, where this information was considered reliable. In general, two years worth of activity data were collected, from April 1999, although illustrative sample data are presented here. The study included every aspect of the emergency care system, from first points of contact (general practice surgeries, NHS Direct, etc) through to points of departure (such as discharge to social service elderly care). Structured interviews were conducted with staff at every stage in the emergency treatment process. There were three objectives of the work:

- (1) To improve understanding of the sources of demand for emergency care within the NHS and the implications for the NHS as part of a supply network.
- (2) To identify how well the NHS responds to demand, using two pilot health regions as research sites for preliminary conclusions.
- (3) To assist in finding methods of changing the design of the treatment process, to improve patient care and effectiveness.

The data gathered were seen to be a snapshot of a broad system, rather than a definitive representation of parts of the system.

RESULTS

When activity levels are studied, in comparison with other service sectors, activity is not especially seasonal, with a 25% difference in activity between the busiest month (June) and the quietest (February). The data are consistent with other years and the different regions exhibited very similar patterns. The seasonal mix variation of patient attendances is more important. Capacity issues are caused by more patient attendances in major illness categories in late autumn and early winter. Neither healthcare region was able to quantify the change in demand for resources attributable to this mix change.

Emergency demand is clearly biased towards patients with minor, non-urgent illnesses. From 121 193 recorded attendances in North Cheshire (two years of data), 85% of annual activity was categorised as non-urgent. If emergency treatment performed in primary care is included, the proportion of minor patients is higher. A significant finding was that in contrast with the perceptions of many hospital staff, the GP network did deal with a substantial emergency workload. Non-urgent attendances at A&E departments were largely attributable to patient self referrals, not instructions or referrals from GPs. For example, in North Cheshire the GP practice that was responsible for the greatest number of weekend out of hours attendances was also providing substantial unused emergency capacity (26 spare appointments each Saturday on average). It was clear that patients were choosing not to attend

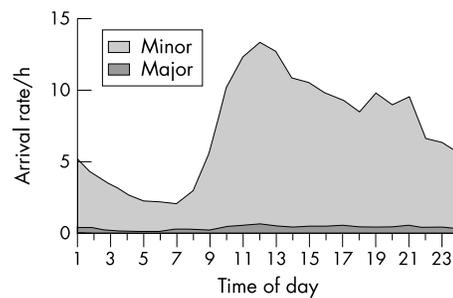


Figure 3 Daily arrival rates for Lewisham A&E patients in January 2001. Source: Adapted from hospital records (5442 attendances).

their GP, despite this service being available. Patients are more consumerist in their attitude to healthcare provision. This potentially has an impact on the relative effectiveness of demand management practices.

The daily fluctuation in demand has a design influence from an operations perspective. Using triage category as a surrogate for type of illness, it is possible to characterise the daily change in demand. Figure 3 shows a typical arrival pattern by hour of day.

Broadly, arrivals of urgent major patients occurred steadily, with the daytime rate being up to twice the arrival rate at night. By contrast, the arrival of minor patients starts usually from 9 00 am and reaches a peak at midday. When this demand is compared with resource availability, it was discovered that nursing resources usually reach a peak at the overlap in shift patterns in the afternoon. Hence the capacity strategy of A&E departments mismatches demand. The situation can be made worse by staff refusing to treat minor patients at night and telling them to return at 9 00 am. This increases the backlog of work in the morning.

The emergency care system needs to be flexible enough to cope with the daily demand changes and it should also have the ability to make adjustments dependent on the type of patient entering the system. All of this has to be achieved within the government’s targets for the speed of response. This makes staff flexibility essential, both in terms of their availability at the predictable peaks of demand and in their ability to cope with the mix of patients. A pre-requisite to achieve fast response is that effective capacity should be in excess of demand.

The pattern of demand showed that there were groups of patients where care pathways could be designed and volume was sufficiently high to make this worthwhile. The manufacturing approach to segmentation divides clusters of work by the similarity of the process sequence and not by other characteristics, for example, reported symptoms. Hence, there are subtle differences between the streaming categories produced by WEST²⁸ and the segmentation used here. This experience is totally consistent with manufacturing industry, where market segmentation does not necessarily correspond with operations segmentation. In the initial analysis, nine segments were defined, but these have since been refined to seven:

- (1) Patients who need advice only (no investigations or treatment)
- (2) Patients receiving a prescription (no investigations)
- (3) Patients receiving minor treatment (no investigations)
- (4) Patients with a minor complaint, requiring investigation
- (5) Patients requiring multiple investigations
- (6) Patients with major illness needing ongoing investigation and treatment
- (7) Patients with major illness involving repeated attendances

The researchers also observed that the triage systems being used by the hospitals create additional problems and use valuable resources in what is a non-value adding task from a

manufacturing perspective. The allocation of priorities beyond a simple “urgent” and “non-urgent” categorisation disrupts the flow of minor patients through the system. Both lead sites had de facto stopped using five categories of triage, knowing that category 5 patients simply do not receive treatment because there is always someone more urgent in the system. Hence, some process segments should not be divided up into subcategories of urgency, except for the most critical.

The process analysis showed that segments 1–5 are normally put into the same minors system, whereby they passed through 5 or 6 stages each entailing unnecessary waiting. Patients who are put through the existing process flow can typically wait for four hours between each stage. Existing practices encourage doctors to assess new, non-urgent patients before they discharge patients who have been treated and are waiting to go home. Patients at the two study sites spent 75% of their time waiting for treatment or discharge. “See and treat” methods of minor patient treatment are a good illustration of a cellular operation. The whole rationale of why this method works is obvious from a manufacturing management perspective. Patients in categories 1–3 do not need to go through unnecessary in-process waits. The see and treat design eliminates non-value adding activity and unnecessary movement of the patient and waiting. The provision of skilled staff in a one stage process design achieves quality and flexibility simultaneously.

A much greater challenge is presented by the design requirements of more complex flows, such as segments 6 and 7 in the above list. These groups of patients need interdepartmental collaboration and multi-agency working. Manufacturing methods would de-centralise support functions, such as radiology and pathology. This would often go further than the simple colocation of treatment and diagnostic support. Some high volume processes would be given their own dedicated facilities of small scale technology, where this exists. This would avoid conflict of priorities, at the expense of apparent utilisation of resources. However, if the technology is small scale, utilisation tends to matter less as a measure of resource efficiency.

Multi-agency work is clearly a future challenge. The research uncovered misunderstanding of the roles of agencies and organisations within the emergency healthcare system. There is also an issue over the different objectives and performance measures applied to these agencies. A balanced, lean healthcare supply system would actually have some parts of the system behaving “sub-optimally” if existing measures and mind-sets are applied. For example, demand-pull control would deliberately generate empty beds, if bottlenecks were seen in social services.

It is clear that the healthcare organisations find internal change less difficult than supply chain improvement. The problem within the emergency system is analogous to Tesco’s experience.¹⁵ By implementing lean thinking, Tesco were able to cut most waiting time from their distribution system. It takes less than five days to buy, distribute, and sell a can of cola. However, the external supply chain cannot match this performance. Each stage of the manufacturing process operates “optimally” according to its own internal efficiency measures. However, this makes the entire production sequence unbalanced. These design incompatibilities mean that it takes in excess of 300 days to make the can from the extraction of raw material. The potential gains in health care are comparable. During the research, the transfer of patients from hospital to community or elderly care was the most obvious example of unnecessary waiting caused by poor coordination.

CONCLUDING REMARKS

Manufacturing theory has now been successfully transferred into many service sector applications. Not only is the application of design principles feasible in health care, manufactur-

ing specialists would see it as necessary to solve some of the queue problems. The results from the research can be generalised, although each site must tailor the methods to suit its own requirements. Modern operations management theories are supportive of what many clinicians have been suggesting for a long time. Firstly, there is now much greater emphasis on quality as an objective. Delays in the system cannot be fully resolved unless the process has the ability to achieve good quality. Healthcare processes have to be accurate, repeatable, and consistent. Secondly, manufacturing theory recognises that speed and flexibility are not encouraged by unhelpful measures of utilisation. The cellular approach champions a decentralised system that permits some resource flexibility in pursuit of its objectives. Thirdly, there is now careful attention paid to task design, so that staff are not subjected to highly repetitive, narrow activities that reduce morale.

These ideas come with their own challenges. In contrast with existing healthcare practices such as triage, cellular segmentation does not trade off performance for one group of patients at the expense of another. Instead, it permits all patients to have the best quality of service that is possible given environmental factors. It will take time for this counter-intuitive argument to be accepted. If staff want an interesting, varied work pattern, a pre-requisite is more development of multi-skilling, especially among nursing staff. Each organisation needs to support the use of nursing staff in more responsible roles and accept that the doctor’s role may involve different, non-traditional tasks. Arguably, the biggest challenge is to implement multi-agency working across the emergency health system. The enormity of this task must be emphasised and current research is actively developing ideas in this area.

REFERENCES

- 1 Quinn JB, Gagnon CE. Will services follow manufacturing into decline? *Harvard Business Review* 1986;**64**:95–105.
- 2 Nishiyama K, Johnson JV. Karoshi—death from overwork: occupational health consequences of Japanese production management. *Int J Health Serv* 1997;**27**:625–41.
- 3 Walley P, Watt A, Davies C, et al. *A study of demand for emergency access health services in two UK health and social care communities*. Warwick: Warwick University report, 2001.
- 4 Hammer M. Reengineering Work: don’t automate, obliterate. *Harvard Business Review* 1990;**68**:104–12.
- 5 Slack N, Chambers S, Johnston R. *Operations management*. 3rd edn. Harlow: Prentice Hall, 2001.
- 6 Hayes R, Wheelwright SC. *Restoring our competitive edge: competing through manufacturing*. New York: Wiley, 1984.
- 7 Schonberger RJ. *Japanese manufacturing techniques: nine hidden lessons in simplicity*. New York: Free Press, 1982.
- 8 Schonberger RJ. *World class manufacturing: the lessons of simplicity applied*. New York: Free Press, 1986.
- 9 Bowen DE, Youngdahl WE. ‘Lean’ service: in defense of a production-line approach. *International Journal of Service Industry Management* 1998;**9**:207–25.
- 10 Krafcik JF. Triumph of the lean production system. *Sloan Management Review* 1998;**30**:41–52.
- 11 Womack J, Jones D, Roos D. *The machine that changed the world*. New York: Rawson Associates, 1990.
- 12 Womack J, Jones D. *Lean thinking: Banish waste and create wealth in your organisation*. New York: Simon and Schuster, 1996.
- 13 Hyer NL, Wemmerlov U. Group technology and productivity. *Harvard Business Review* 1984;**62**:140–50.
- 14 Parnaby J. A systems approach to the implementation of JIT methodologies in Lucas Industries. *International Journal of Production Research* 1988;**26**:483–92.
- 15 Swamidass PM. *Technology on the factory floor II*. Washington DC: The Manufacturing Institute, 1994.
- 16 Kher HV, Jensen JB. Performance improvement and process focus: a case for hybrid cellular manufacturing. In: Blackmon K, Brown S, Cousins P, et al, eds. *What really matters in operations management*. Bath: Euroma Annual Conference, 2001:947–56.
- 17 Skinner W. The focused factory. *Harvard Business Review* 1974;**52**:113–21.
- 18 Levitt T. Production-line approach to service. *Harvard Business Review* 1972;**50**:20–30.
- 19 Levitt T. The industrialization of service. *Harvard Business Review* 1976;**54**:32–43.
- 20 Taylor S, Lyon P. Paradigm lost: the rise and fall of McDonaldization. *International Journal of Contemporary Hospitality Management* 1995;**7**:64–8.

- 21 **Pine BJ.** *Mass customization: the new frontier in business competition.* Boston, MA: Harvard Business School Press, 1993.
- 22 **Silvestro R.** Positioning services along the volume-variety diagonal. The contingencies of service design, control and improvement. *International Journal of Operations and Production Management* 1999;**19**:399–420.
- 23 **Heskett JL.** *Shouldice Hospital Limited, Case 9–683–068.* Boston, MA: Harvard Business School, 1993.
- 24 **Gummesson E.** Are you looking forward to your surgery? *Managing Service Quality* 2001;**11**:7–9.
- 25 **Vaughan RW,** Aluse JJ. Factory focus in hospital-owned ambulatory surgery. *International Journal of Service Industry Management* 1992;**3**:1992:63–75.
- 26 **Bushell S,** Shelest B. Discovering lean thinking at progressive healthcare. *Journal for Quality and Participation* 2002;**25**:20–5.
- 27 **Heinbuch SE.** A case of successful technology transfer to healthcare: total quality material management and just-in-time", *Journal of Management in Medicine* 1995;**9**:48–56.
- 28 **WEST.** *Reforming emergency care.* London: Department of Health, 2001.