Failure demand: An evaluation of concept in UK Primary Care

Author 1: Paul Walley, PhD, Senior Lecturer, The Faculty of Business and Law, The Open University Business School, The Open University, Milton Keynes, UK, Telephone: 01908 654992, e-mail: paul.walley@open.ac.uk

Author 2: Pauline Found**,** PhD**,** Professor, Business School**,** University of Buckingham**,** Buckingham, UK, Telephone: 01280 814080, e-mail: Pauline.found@buckingham.ac.uk

Author 3: Sharon J. Williams**,** PhD, Senior Lecturer, College of Human and Health Sciences, Swansea University, Swansea, UK, Telephone: 01792 295662, e-mail: Sharon.j.williams@swansea.ac.uk

Corresponding author: Sharon Williams
Corresponding Author's E-mail: Sharon.j.williams@swansea.ac.uk

Abstract

Purpose: Our purpose was to assess failure demand as a lean concept that assists in waste analysis during quality improvement activity. We assess whether the concept's limited use is a missed opportunity to help understand improvement priorities, given that a UK Government requirement for public service managers to report failure demand has been removed.

Design/methodology/approach: We look at the literature across the public sector and then apply the failure demand concept to the UK's primary healthcare system. The UK National Health Service (NHS) demand data are analysed and the impact on patient care is elicited from patient interviews.

Findings: The study highlighted the concept's value, showing how primary care systems often generate failure demand partly owing to existing demand and capacity management practices. This demand is deflected to other systems, such as the accident and emergency department, with a considerable detrimental impact on patient experience.

Research implications: More research is needed to fully understand how best to exploit the failure demand concept within wider healthcare as there are many potential barriers to its appropriate and successful application.

Practical implications: We highlight three practical barriers to using failure demand: (i) poor general understanding of demand within the healthcare system; (ii) limited understanding of systems improvement; and (iii) need to apply the concept for improvement and not just for reporting purposes.

Originality/value: We provide an objective and independent insight into failure demand that has not previously been seen in the academic literature, specifically in relation to primary healthcare.

Keywords: Primary healthcare, Failure demand, Demand, Capacity, Patient satisfaction

Article Classification: Case study

Introduction

All public service sector managers are under significant financial pressure, especially after the 2007 global financial crisis. The UK Government spending on public services and welfare in 2015 was £747bn, but there have been several financial cuts to public expenditure, with the 2013 financial review requiring £11bn more savings in the 2015/16 budget (ONS, 2016). The challenge for service managers is to identify ways to reduce costs without unnecessarily cutting service quantity and quality. Many authors looked towards lean or quality improvement approaches to achieve sustainable productivity gains without compromising services. Attempts to implement such approaches have been met with considerable difficulties but some significant improvements have been achieved in various public-sector organisations (McNulty, 2003; Lucey *et al.*, 2005; Radnor and Osborne, 2013).

One approach to implementing lean or systems thinking in the UK public sector is the Vanguard Method (Seddon, 2003), which focuses on service purpose, in customer terms, and pays attention to service demand type and frequency. Using a systems perspective, the method then studies system capability to meet purpose and demand. Like many approaches, there is a focus on making work flow through the system and eliminate waste. Vanguard case studies about the system have been published (Seddon, 2003, 2008; Middleton, 2010). The OECD, in a system review approach to public service improvement, pick out the Vanguard Method as one that has led to process innovation and change (Cook and Tonurist, 2012). The Vanguard method is a systems approach to services redesign to improve services and efficiency.

One idea contained within the Vanguard Method is *failure demand*, defined as 'demand caused by a failure to do something or do something right for the customer' (Seddon, 2003, p26). In a healthcare context, failure demand might appear in many forms, ranging from unnecessary prescriptions or repeated diagnostic tests, to repeat patient presentation owing to failure to treat a condition at first contact. Other activities, such as unnecessary follow-up visits, might also be classed as failure demand. When discussing service improvement, Seddon (2009, p.33) makes a bold claim: 'In service organisations … failure demand often represents the greatest lever for performance improvement. In financial services, it can account for anything from 20 to 60 per cent of all customer demand … in local authorities and police forces as much as 80-90 are avoidable and unnecessary'. It is surprising that failure demand has not received more attention, especially its merits and investigation into its use. We seek, therefore, to contribute to knowledge in several ways. First, we show how failure demand analyses might be conducted and how this could be used to act as a driver for system improvement, based upon existing literature. We critique the UK Government's failed attempt to implement failure demand (adapted and termed avoidable demand) as one performance measure that public service managers had to submit in their performance reports. Second, we investigate how failure demand might be categorised within the UK National Health Service (NHS), focusing on how primary care demand is managed and the impact on work in the healthcare. We also include qualitative interviews, which examine failure demand's impact on patients and their service experiences.

Healthcare context

Projected demographic and societal changes are expected to intensify the pressure on health systems and demand new and improved healthcare. Ageing populations in both emerging and developed nations are increasing healthcare demand. According to the United Nations, the world's population is expected to increase by one billion people by 2025 (United Nations Population Fund 2013); 300 million will be people aged 65 or older, as globally, life expectancy continues to rise. Healthcare resources and service innovation are needed to support this increase. Countries will be affected differently by these demographic changes. Successful and sustainable change will require flexible and adaptive models to fit the new health economies.

Like many other healthcare systems world-wide, the UK's NHS is under pressure from rising deficits, worsening performance and declining staff morale, which means that the NHS, a public healthcare system free at the point of use, is facing its biggest challenge for many years. The NHS Five Year Forward View (NHS England, 2014) estimates that the NHS needs additional £8 billion a year by 2020/21, and this projection depends on delivering £22 billion efficiency savings a year by 2020/21; described as ambitious and which requires much higher productivity improvements than NHS staff have historically been able to deliver. The UK governments have relied on external pressures such as targets, inspection and competition to drive reform and improvements (Audit Commission, 1999). This approach has delivered mixed results. Examining some high-performing health systems in the UK and internationally, suggests that organisation staff, in their pursuit to transform care, should ensure leadership, engaging staff and focusing on a clear commitment to put patients first (The King's Fund, 2015). In 2016 NHS England staff took steps to relax the key waiting time targets for more than 50 English hospitals to help ease their financial problems (NHS England, 2016). The UK think tank, the King's Fund, called for greater emphasis on how services need to change, the starting point being new care models proposed in the NHS Five Year Forward View. A recent five-year plan review mentions the need for front-door clinical streaming, improving patient flow and managing avoidable demand (NHS England, 2017).

Here we focus on NHS primary care rather than all healthcare services (e.g., acute care). Primary care is often the first contact for people needing healthcare, which is provided by professionals such as general practitioners (GPs), dentists and pharmacists. The primary care system is responsible for 80% of all contact with NHS patients. Sixty-two percent of primary care contact is with a local GP and the rest is mainly with practice nurses. Eighty-two percent of contact is at practice premises, and an additional 12% by telephone. One issue that the NHS must address is what happens when patients require access to primary care services when GP facilities are unavailable. 'Out-of-hours' (OoH)is usually defined as the period 6.30 pm to 8.00 am. The responsibility for this service depends on whether GPs opted out of providing OoH care, but in many cases the responsibility now lies with Care Commissioning Groups (CCGs), where some or all GPs within a region opted out. English CCGs' original purpose was to give GPs a greater input into how services are designed and provided within the community. This led, for example, to the growth in alternative treatment provision, such as out-of-hours surgeries, walk-in centres and urgent care centres (Roland *et al.,* 2012).

There have been concerns expressed recently about the primary care system's ability to cope with demand. Dayan *et al.,* (2014) debated the long-term prospects for primary care in the UK and highlighted that GPs have experienced a 20% loss in real terms income in recent years and there are longer-term recruitment pressures owing to impending retirements and recruitment shortages. A recent GP Patient Survey (Ipsos MORI, 2017) reported 32 per cent of patients found it difficult to get through on the phone to their surgery, and 16% were unable to get an appointment to see or speak to someone the last time they tried. Nearly one in five patients (18%) said they tried to contact an NHS service in the past six months when they wanted to see a GP, but the surgery was closed, either for themselves or for someone else.

There is some concern that increasing attendances per patient has put pressure on patients a GP is able to care for at any time, resulting in an overload on the requests for appointments. In the UK, GPs care for typically between 1,700 and 2,000 patients per full time equivalent, but there is variation in these figures (HSCIC, 2016). In the United States, patient to GP ratio is referred to as 'panel size' and there are debates about the correct proportion, partly influenced by demographic factors (Murray *et al.,* 2007). There are other reasons why access can be limited, including issues with the way GPs manage patient schedules and the difficulties associated with recruiting GPs in the UK. Kiran and O'Brien (2015, p.399), drawing on the work of Pope *et al.,* (2008), commented: Many primary care practices in England

misinterpreted advanced access, often embargoing '70% of appointment slots for same-day appointments, and then requiring patients to call immediately after the office opened to book into these embargoed slots … this resulted in frustrated patients who complained that the new system made it impossible to book appointments with their doctors in advance. This approach has poignantly been referred to as 'access by denial'.'

In the winter, 2014, there were considerable issues reported regarding access to UK primary care and worse problems associated with long queues and delays in Accident and Emergency (A&E) departments, with near gridlocked patient flows across the system (Blunt *et al.,* 2015). Patients frequently find there were no available bookable slots for appointments, or that waiting times for non-urgent appointments were extending into many weeks' wait. Many GPs keep appointments in reserve as same-day, urgent slots, resulting in long queues waiting at surgeries or phoning through early in the morning to gain a same-day appointment. Patients who called later in the day were often told to try again the next day as all appointments were taken. Similar issues were experienced with accessing emergency services in 2015 and again in 2016. Many patients unable to obtain care from their GP will try to seek access to other healthcare (NAO, 2015).

Managing capacity and demand is not well-versed within healthcare. Research focuses on secondary care (e.g., hospitals) waiting times and waiting lists management with many focusing on how demand and capacity are measured (Silvester *et al.,* 2004). For example, Westbury *et al.,* (2009) examine how demand and capacity are measured within a discrete surgical speciality, leading to an ability to perform sophisticated analyses and achieve further improvements. Walley (2013) reports public services as 'resource-driven' rather than 'demand-driven' when planning resources to meet demand. This approach has detrimental effects on the way resources are allocated, especially in the medium term. The practical effect is failure to understand demand patterns and how to manage them, creating unnecessary queues or wasting resources. The general recommendation is, therefore, that pubic service managers should adopt private sector capacity planning practices.

Failure demand

Failure demand emerged from a telephone call study suggesting that most calls received by a sales call centre were, in practice, calls from customers complaining about the service they had received or reporting issues for the company to deal with, not the desired calls from customers wishing to place orders. Marr and Neely (2004, p.4) argue 'It is critical to understand and classify the nature of demand. Often calls are unwanted or even unwittingly generated by other parts of the organization. By analyzing and classifying demand, it becomes more manageable and more predictable. Unwanted calls could then be reduced or even eliminated.' Thus, in lean thinking terms, these calls indicate waste in the system and handling the call is a waste. A key mistake made by managers was to treat all work coming into the system as demand, rather than seeing an opportunity to reduce workload on staff and simultaneously improve customer service. These analyses have since been applied across the UK public sector, with many reports coming from local council services such as (Randle and Kippin, 2014). Housing services are the most widely reported applications, with examples provided from several sources (ODPM, 2005; Jackson, *et al.,* 2007; Masters, 2009; Zokaei *et al.,* 2010). Figure 1 is adapted from data provided by Seddon (2003), based on Jackson *et al.,* (2007), breaking down calls coming into a council housing department call centre.

Figure 1 here

Figure 1 calls are shown as failure demand (darker) and true demand (lighter); the latter includes only new demand that has not been requested previously. So, a first call to request a

plumber is true demand, whereas a follow-up call to enquire when the plumber is arriving is failure demand. This analysis is appealing, as it holistically provides a true insight into system waste and what is creating that waste. Anyone familiar with traditional lean thinking tools would also see the similarity to conventional Pareto analysis.

Failure demand measurement in government
In 2008, following a pilot study, local government managers were asked to routinely measure and report their failure demand. This announcement had been preceded by prior reports into contact between public services and the public (Masters, 2009). For example, one underperforming local council had been identified as logging two-thirds of all contact as waste demand (Caulkin, 2005). Benefits service errors had also been wasteful, where welfare under- and over-payments were resulting in considerable unnecessary contact. The cabinet office head, Alexis Cleveland, announced that National Indicator 14 (NI 14) would be Avoidable Contact, as the Government's approach to measurement and recording failure demand. Measurement started in October 2008 for first reporting in April 2009. The Government advice defined avoidable contact in several ways:

1. Unnecessary clarification sought by the customer; a phone call or email asking to explain a poorly-worded communication.
2. Incorrect contact; the customer is passed to the wrong department or transferred to the wrong number.
3. Repeat contact; the customer must contact the organisation multiple times for the same reason; e.g., having to report an address change to several departments.
4. Customers having to progress-chase work, including enquiries why a home appointment had been missed or pre-arranged refuse collection had not occurred.
5. Repeat contact owing to unfinished work, including closing a job request before the work had been completed to the customer's satisfaction (IDEA, 2008).

The measure was controversial. As one report stated: 'The underlying aim of NI14 is laudable … However, vociferous debate … has been polarised between those who believe that NI14 will be an important tool for driving transformation and aligning efficiency … and those who believe that NI14 will hinder the process of transformation and divert scarce resources to monitoring activities.' (LGITU, 2008, p.3). Concerns were raised about implementing the measure effectively. The LGITU study (ibid) lists potential barriers to its implementation:

- Silo mentality (not sharing knowledge or information).
- Staff training costs.
- Weak guidance.
- Poor support from senior stakeholders.
- Capacity/skill sets internally.
- Technology infrastructure.
- Conflicting policy priorities.
- Poorly joined up services.
- Proprietary/incompatible systems.
- Analysis costs.
- Data collection costs.
- No budget to implement changes.

These NI14 concerns were probably justified owing to the measure's longevity. In March 2010, it was discreetly removed from the Government's measures (Martin, 2010). The main reasons

publicly given were technical issues about data collection and reporting costs. It was also evident that some council staff were reluctant to report the measure accurately and there were concerns about the value that councils were getting out of the measurement.

Failure demand in primary care
In healthcare, the issue is whether patients can receive the right care in the right location without unnecessary delay. Failure demand, should it exist, may include healthcare system contact that is either unnecessary extra steps in the process to obtain care, or steps in the process that do not provide appropriate care. Comparable studies on access issues use two other terms: supplier-induced demand (Bickerdyke *et al.,* 2002) and supply-sensitive demand (Dartmouth, 2007) (Table I). In an integrative study, Rosen (2014) found information that suggests 16% of patients attending walk-in clinics would not have bothered to seek any care if this service had not been available. Additionally, 46% using walk-in centres should not have needed to attend before obtaining the care they needed, which added an extra step in their journey and created significant extra health service demand/contact. In some cases, no treatment was needed but in others, treatment could have been obtained more directly; e.g., visiting a pharmacy. The relative proportions are not recorded. Appleby (2013) showed that substitutes for emergency visits to primary care or A&E departments, such as walk-in centres, minor injuries units and urgent care centres, made no difference to A&E attendance rates once they were implemented; in effect, additional demand.

Table I here

Formal reports mentioning failure demand in healthcare are uncommon, but small studies indicate the scope and potential for further study. One report (Locality, 2014) included two observations that highlight systemic failure demand generated within the healthcare system: 'Studying 21 people with health needs revealed that they created 79 demands on the acute healthcare system, 75 demands into GPs, 55 demands on district nurses and 30 demands on adult social care. Another study analysed eight peoples' records going back between one and nine years. Collectively, these eight individuals exited and re-entered the system 124 times, and were subjected to 236 'assess–do–refer' cycles.' Locality (2014, pp. 16-17). This work appears to support the potential for more healthcare failure demand analysis, with the possibility that failure demand is considerable across the healthcare system. In our opinion, unmet demand within primary care potentially affects the remaining system; e.g., it is suggested that unmet primary care demand can flow to the emergency care system, which, in turn, can slow down the emergency care system as it struggles to deal with higher demand (Blunt *et al.,* 2015).

Research methods
We combine two studies. First, in March 2015, an expert panel considered the NHS 2014/15 'Winter Crisis'. Ninety healthcare management experts convened for a round table discussion, studying each healthcare system element and its role in winter pressures. The expert panel was chosen from senior managers, healthcare professionals, technical specialists and improvement consultants from across NHS organisations (including Scotland and Wales), civil service, universities, independent research organisations and private consultancy companies. Organisations such as the Nuffield Trust, the Emergency Care Intensive Support Team and Monitor were included in the event. The panel was initially split into groups of ten, with care taken to ensure diverse specialities represented at each table. After an introduction where the Winter Crisis was described in detail, the remaining time was structured into three sessions, where an issue was highlighted, and each table was given the opportunity to discuss and

recording the issues and reporting their findings in a plenary session. These responses were summarised by appointed chairs at each table, recorded in word documents, and handed over at the end. Other points by individuals were captured on notice boards, which were photographed. Discussants were asked to highlight evidence sources for points raised and this resulted in 22 follow-up responses with data after the event. The following questions were posed:

1. What drives A&E pressures?
2. What solutions have worked and failed?
3. Can we model the system to understand change?

The first session included demand in the system and potential failure demand created by changes to system capacity over the winter period. To understand the problems, experts were asked to provide data on demand patterns in their system. Follow-up work meta-analysed existing data and generated demand flow case material and capacity provision across NHS England and Scotland, which was used to look at primary care demand, patient flows and impact on service quality. We also consider the impact on other services such as the NHS 111, which is the 24-hour telephone service for patients/relatives when there is an urgent healthcare need, but not a 999-emergency call requiring an emergency service response. The service is provided by trained advisers who are supported by healthcare professionals. During the call, patients/relatives are asked questions about their symptoms before redirecting to the best medical care provider. Failure demand issues were also elicited from patient interviews, which focused on patient experiences when accessing primary care and managing interventions/interactions by/with various community services. Fifteen exploratory interviews were conducted with respiratory and cardiac patients who were managing the complexities associated with long term conditions. The interviews were audio recorded and transcribed. The data were analysed using a thematic analysis template (King, 2004). This iterative approach develops conceptual themes. Where appropriate, quotations were included within the findings.

Findings

We found that demand data are not routinely collected, and it is unclear how much unmet primary care demand was not met. Individual primary care systems were not in place to capture all potential demand during attempts to access the system; e.g., staff in many practices do not monitor or record failed calls when requested services were unavailable or appointments rejected by callers for reasons such as the delay between call and appointment offered. Staff record activity; i.e. work done, rather than demand. Statistics showed how demand is deflected to alternative contact points, especially the 111 services, where there are clear, seasonal out-of-hours contact patterns, which are exaggerated during holiday periods (Figure 2). These patterns coincide with reduced primary care system capacity; i.e., the conventional working week is reduced by public holidays.

Figure 2, 3 and 4 here

Data show that between 45-55% of calls result in a recommendation to the patient that they attend primary care when non-availability was probably the reason for the call to 111. Primary care's effect on the 111 service is identifiable; over the Christmas period, service staff struggle to cope with the increased demand at a time when they have their own staffing challenges, which results in more abandoned calls (Figure). What happens to demand that cannot be met because the primary care is closed when demand occurs – NHS 111 service staff have no choice

but to pass demand onto emergency care departments if triage suggests patients need medical attention. The NHS A&E statistics show how demand patterns emerge (NHS, 2015).

Impact on patient care
Interviewees were reluctant to seek assistance from emergency care service staff. If their condition and symptoms allowed, all interviewees preferred to seek assistance from their GPs. As patients became more skilled at managing their conditions they reported they were more likely to seek assistance early when they could access their GP surgeries or other primary care providers, such as specialist community nurses. They reported that advice could be via a telephone rather than a face-to-face. The interviewees recalled different appointment systems operated by their GP practice staff, with some offering open surgeries (where no appointment times are given) but patients are seen based on arrival times. Others mentioned the frustrations of ringing for appointments during the morning, often being held in a queue before being informed that no appointments were available, and to call back later that day or the next morning. One interviewee recalled during the early stages of his respiratory condition, that he did not seek medical attention until an acerbation had occurred, which led him to attend his local emergency department, which led him to hospital admission. He mentioned that his condition worsened in the evenings when he was less likely to have access to primary care services. Another patient who became unwell after being discharged after heart surgery resulted in an attendance to his local emergency department (on the NHS helpline advisor's guidance). The patient was not admitted but medication was prescribed. From regular follow-up appointments with his GP, some four days later, it was evident that his visit to the emergency department was not registered with his GP practice. Hence, the difficulties associated with detecting and recording failure demand across healthcare providers.

Discussion
One challenge faced by the researchers was to obtain validated primary care demand data at practice level. How individual practice staff manage demand means that determining true demand on the system is difficult. General practitioners might not understand how much unmet demand there is because patient demand is not always recorded. It is not clear from our analysis what proportion simply abandon their attempts to obtain medical care when they experience problems. Similarly, the way in which GPs attempted to implement *advanced access* inadvertently converts routine demand into emergency demand because routine demand cannot reliably be booked in a reasonable timescale. Patients requiring an appointment within one week may have to convert this demand into an emergency request. This creates failure demand in the emergency care system. Our analysis shows additional demand does not increase hospital admissions. Patients unable to access primary care are discharged from the system prior to full admission. However, there will be an increase (up to 50%) in A&E minor patient (e.g., less severe injuries) workload and an unknown increase in patients triaged as A&E 'majors'; i.e., on arrival at A&E, patients are assessed to ensure that people with the most serious conditions and life-threatening emergencies, are seen first (NHS Choices, 2017).

Patients are generally reluctant to access emergency services when primary care services are lacking. Some patients mentioned that, as they became more confident in managing their conditions and are less dependent on their GP and other services. They were more able to identify exacerbations or early infections and therefore, accessed appropriate services during 'normal' opening times. The expert patient programme in the UK is designed to help improve patient outcomes. For example, there are formal (often education-based) patient expert programmes, which are designed to help patients to manage their condition (Griffiths *et al.,* 2007). There were occasions, after seeking advice from national NHS helplines, such as 111, where patients were advised to attend their local emergency departments or services.

Sometimes patients sought advice from specialist community nurses, but these seemed to be different across the commissioning landscape, which can be problematic, when patients aren't sure about the care they need, or whether a service is available.

Data strongly suggest that requests recorded within the UK healthcare system can be classed as failure demand and there would be a significant reduction in waste and an improvement in patient experience if the failure demand is reduced. We question why the NI 14 measure was removed, and hence why there is currently no pressure to use failure demand within the healthcare system. Our literature search and follow-up analysis identify three debates:

1. Should demand measurement be continuous?
Seddon (2009) recommends that demand and failure demand measurement should be a snapshot. The overall picture that the failure demand analysis provides is normally a system failure through errors and poor system design. He argues that the underlying NI 14 measure issue was continuous measurement and the behavioural consequences (in a 'command-and-control' system), which leads to wrong behaviour. One default command-and-control behaviour is to develop systems to manage reporting, not address the underlying problems, which become a 'tick box' exercise. Seddon's recommendation contrasts with the ODPM (2005) report, which encourages the Vanguard method. The ODPM report puts failure demand data into a continuous statistical process control (SPC) chart, which shows the dilemma faced by users. Although it is expected that failure demand reduces over time as systems improve – and the SPC chart demonstrates this system improvement - there are problems associated with continually reporting failure demand. It becomes too tempting to turn the failure demand measure into a target, where managers intervene when failure demand temporarily increases. This continual measurement also misses the point about failure demand information's diagnostic value. If the focus is on how the measure can be collected and reported efficiently – without using the data to help improve the system, then the purpose of the measure has been missed.

We could argue that, although demand for healthcare is relatively stable, there are periods when demand changes through epidemics or other special events, or because decision-making by managers; e.g., changes to ambulance routes and case-mix variation, especially monitoring increases in elderly care workload, are also crucial factors that affect the system's ability to cope with demand, which needs to be tracked so that changes are identified. Where the mechanisms we have identified are considered, changes in actual demand also affects failure demand, and so increases in failure demand may be indicate that a system out of control. In UK primary care, failure to understand demands placed on the system must be addressed; otherwise appropriate capacity in the system cannot be understood.

2. Is failure demand addressed in the right way?
We made an analogy between failure demand and Pareto analysis in other lean or total quality management settings. We suggest that one reasons for limited success in the public sector is the linkage between analysis and system behaviour. When a conventional approach to process improvement is matched with a failure demand analysis, there may often be a gap in the root cause analysis; i.e., why the failure demand occurs. In conventional Pareto analysis, diagnosis would be accompanied by system assessment such as a '5 whys' assessment. Managers without the appropriate systems training tackle the symptoms, not address the system issues (Trbovich, 2014). Failure demand must be linked to full systems analysis (including appropriate systems mapping).

3. Is failure demand measurement used for the right purpose?

Seddon (2009) said that using failure demand to set targets misses its purpose. Our experiences suggest that the using failure demand as a performance measure is clearly inappropriate and is strictly a diagnostic and indicative tool.

Conclusions
We open the failure demand dialogue in relation to accessing healthcare emergency services. We draw upon UK national demand data and patient experience interviews to convey how failure demand might present in healthcare. Our overall aim was to assess failure demand as a lean concept to analyse waste during quality improvement activity. Our contribution is three-fold: first, providing insight in to how failure demand analysis may be conducted, Second, showing how failure demand can affect the healthcare system. Third, indicating how failure demand affects patients. We confirm that limiting failure demand analysis is a missed opportunity to help understand healthcare improvement priorities. We recognise the difficulties associated with recording failure demand activity particularly across healthcare providers. We focused mainly on NHS primary care. Further research is required, therefore, to explore failure demand in other healthcare settings, such as hospital departments. Our research has implications for academics, practitioners and policy makers. For academics, it is a call to generate further empirical studies to broaden our failure demand understanding, particularly within healthcare. For practitioners, there is a need to identify failure demand when looking to redesign and improve services. For policy makers, there is a need to support managers trying to measure and reducing failure demand within their services.

References
Appleby, J. (2013), 'Pressures on accident and emergency services', http://www.kingsfund.org.uk/audio-video/john-applebypressures-accident-and-emergency-services, accessed June 2017.
Audit Commission (1999), *Performance Measurement as a Tool for Modernising Government*, Audit Commission, London
Bickerdyke, I. Dolamore, R. Monday, I. and Preston, R. (2002), *Supply Induced Demand for Medical Services,* Productivity Commission, Melbourne, Australia.
Blunt, I. Edwards, N. and Merry, L. (2015), *What's behind the winter crisis?* Nuffield Trust, London.
Caulkin, S., (2005), 'How to be big and beautiful', The Observer, 30 Jan, https://www.theguardian.com/society/2005/jan/30/publicservices.business, (accessed Jan 2018.
Cook, J. and Tonurist, P., (2012), *From Transactional to Strategic: systems approaches to public service challenges*, The observatory of Public sector Innovation, OECD, Paris.
Dartmouth Atlas Project Topic Briefing. (2007), 'Supply Sensitive Care', <http://www.dartmouthatlas.org/downloads/reports/supply_sensitive.pdf>, accessed May 2017.
Dayan, M., Arora, S., Rosen, R. and Curry, N. (2014), *Is General Practice in Crisis?* Policy Briefing#1, November, Nuffield Trust, London.
Griffiths, C., Foster, G. Ramsay, J., Eldridge, S. and Taylor, S. (2007). 'How effective are expert patient (lay led) education programmes for chronic disease?', *British Medical Journal*, Vol. 334, pp. 1254-1256.
HSCIC. (2016), *General Practice Trends in the UK to 2015*, Health and Social Care Information Centre, London.
IDEA. (2008), *Reducing Avoidable Contact, a guide to N14*, Improvement and Development Agency, London.
Ipsos MORI. (2017), *GP Patient Survey: National summary report*, Ipsos Mori, London.

Jackson, M., Johnston, N. and Seddon, J. (2007), 'Evaluating Systems Thinking in Housing', *Journal of the Operational Research Society,* Vol. 59, pp. 186–197.

King, N. (2004), *Using templates in the thematic analysis of text*, in C. Cassell, & G. Symon, (Eds), Essential Guide to Qualitative Methods in Organisational Research, Sage Publications, London.

Kiran, T. and O'Brien, P. (2015), 'Challenge of same-day access in primary care', *Canadian Family Physician,* Vol. 61 No. 5, pp. 399-400.

LGITU. (2008), *NI14: Help or Hindrance?* Informed Publications, London.

Locality. (2014), *Saving money by doing the right thing*, Locality, London.

Lucey, J., Bateman, N. and Hines, P. (2005) 'Why major Lean transitions have not been sustained',. *Management Services*, Vol. 49, Summer, pp. 9–13

Marr, B. and Neely, A. (2004), Managing and Measuring for Value: The Case of Call Centre Performance, *Cranfield School of Management Report*, Cranfield, UK.

Martin, P. (2010), 'Avoidable Contact: If only…', <http://www.lgcplus.com/news/finance/efficiency/avoidable-contact-if-only/5013769.article>, accessed June 2017.

McNulty, T. (2003), 'Redesigning Public Services: Challenges of Practice for Policy', *British Journal of Management*, Vol. 14 No. S1, pp. 31–45

Masters, K.I.A. (2009), *From New Public Management to Lean thinking: understanding and managing potentially avoidable failure induced demand*, Unpublished PhD Thesis, University of Surrey, UK.

Middleton. P, (ed.). (2010), *Delivering public services that work (Volume 1): systems thinking in the public sector—case studies*, Triarchy Press, Axminster

Murray, M., Davies, M. and Boushon, B. (2007), 'Panel size: How many patients can one doctor manage?', *Family Practice Management*, Vol. 4 No. 4, pp. 44-51.

NAO. (2015), *Stocktake of access to general practice in England*, National Audit Office, November, London.

NHS Choices. (2017), Urgent and emergency care services in England, https://www.nhs.uk/NHSEngland/AboutNHSservices/Emergencyandurgentcareservices/Pages/AE.aspx, accessed January 2018.

NHS England. (2014), 'Five year forward view', <https://www.england.nhs.uk/wp-content/uploads/2014/10/5yfv-web.pdf>, accessed June 2017.

NHS England. (2015), 'A&E attendances and emergency admissions', https://www.england.nhs.uk/statistics/statistical-work-areas/ae-waiting-times-and-activity/statistical-work-areasae-waiting-times-and-activityweekly-ae-sitreps-2015-16/, accessed January 2018.

NHS Wales. (2016), 'Written statement: NHS Reforms', <http://www.wales.nhs.uk/newyddion/13330>, accessed July 2017.

NHS England. (2017), 'Next steps on the NHS five year forward view' <https://www.england.nhs.uk/wp-content/uploads/2017/03/NEXT-STEPS-ON-THE-NHS-FIVE-YEAR-FORWARD-VIEW.pdf>, accessed August 2017.

ODPM. (2005), *A Systematic Approach to Service Improvement*, September, ODPM Publications, London.

Olsen, H. (2008), 'NI 14: Help or Hindrance?' <http://www.ukauthority.com/About-Us/Research-Projects/NI14.aspx>, accessed June 2017.

ONS. (2016), 'UK Perspectives 2016: Spending on public services in the UK', <http://visual.ons.gov.uk/uk-perspectives-2016-spending-on-public-services-in-the-uk/>, accessed July 2017.

Pope, C., Banks, J., Salisbury, C. and Lattimer, V. (2008), 'Improving access to primary care: eight case studies of introducing advanced access in England', *Journal of Health Services Research Policy,* Vol. 13 No. 1, pp. 33-39.

Radnor, Z. and Osborne, S.P. (2013), 'Lean: A failed theory for public services?', Public Management Review, Vol. 15 No. 2, pp265-287

Randle, A. and Kippin, H. (2014), *Managing Demand: Building future public services*, RSA, London.

Rosen, R. (2014), *Meeting Need or Fuelling Demand, Improved access to Primary care and Supply-Induced Demand*, Nuffield Trust, London.

Roland, M., Guthrie, B. and Thomé, D.C. (2012), 'Primary medical care in the United Kingdom,' *Journal of American Board Family Medicine*, Vol. 25 No. 1, pp. S6-S11.

Seddon, J. (2003), *Freedom from Command and Control*, Vanguard Press, Buckingham.

Seddon J. (2008), *Systems thinking and the public sector*, Triarchy, Axminster.

Seddon, J. (2009), 'Failure Demand – from the horse's mouth', *Customer Strategy*, Vol. 2 No. 1, pp. 33-34.

Silvester, K., Lendon, R., Bevan, H., Steyn, R. and Walley, P. (2004), 'Reducing waiting times in the NHS: is lack of capacity the problem,' *Clinicians in Management*, Vol. 12, pp. 105-111.

The King's Fund. (2015), 'Health and social care: Three priorities for the new government', <http://www.kingsfund.org.uk/projects/new-gov/three-priorities-new-government>, accessed May 2017.

Trbovich, P. (2014), 'Five ways to incorporate systems thinking into healthcare organisations', *Horizons*, Vol 31, No. 4, , pp. 31-36.

United Nations Population Fund. (2013). 'World population to increase by one billion by 2025', <http://www.unfpa.org/news/world-populationincrease-one-billion-2025>, accessed May 2017.

Walley, P. (2013), 'Does the public sector need a more demand-driven approach to capacity management?', *Production Planning and Control*, Vol. 24 No. 10-11, pp. 877-890.

Westbury, S., Pandit, M. and Pandit, J. (2009), 'Matching surgical operating capacity to demand using estimates of operating times', *Journal of Health Organisation and Management*, Vol. 23, No. 5, pp. 554-567.

Zokaei, Z., Elias, S., O'Donovan, B., Samuel, D., Evans, B. and Goodfellow, J. (2010), '*Lean and Systems Thinking in the Public Sector in Wales*', Lean Enterprise Research Centre report for the Wales Audit Office, Cardiff University, Cardiff.

Table I: Demand types in healthcare

| Demand type	Description
Supplier-induced demand	Demand created by providers as imperfect care agents
Supply Sensitive demand	Increased health service use stimulated by increased (regional) supply
Induced demand	Attendance at walk-in centres or 111 services where the patient would not have used a service if it was unavailable
Induced utilisation	Service use following an initial contact with a direct access service, which would have occurred even if the direct access service was not available and are therefore additional attendances

Figure 1: Failure demand analysis (Adapted from Seddon, 2003)
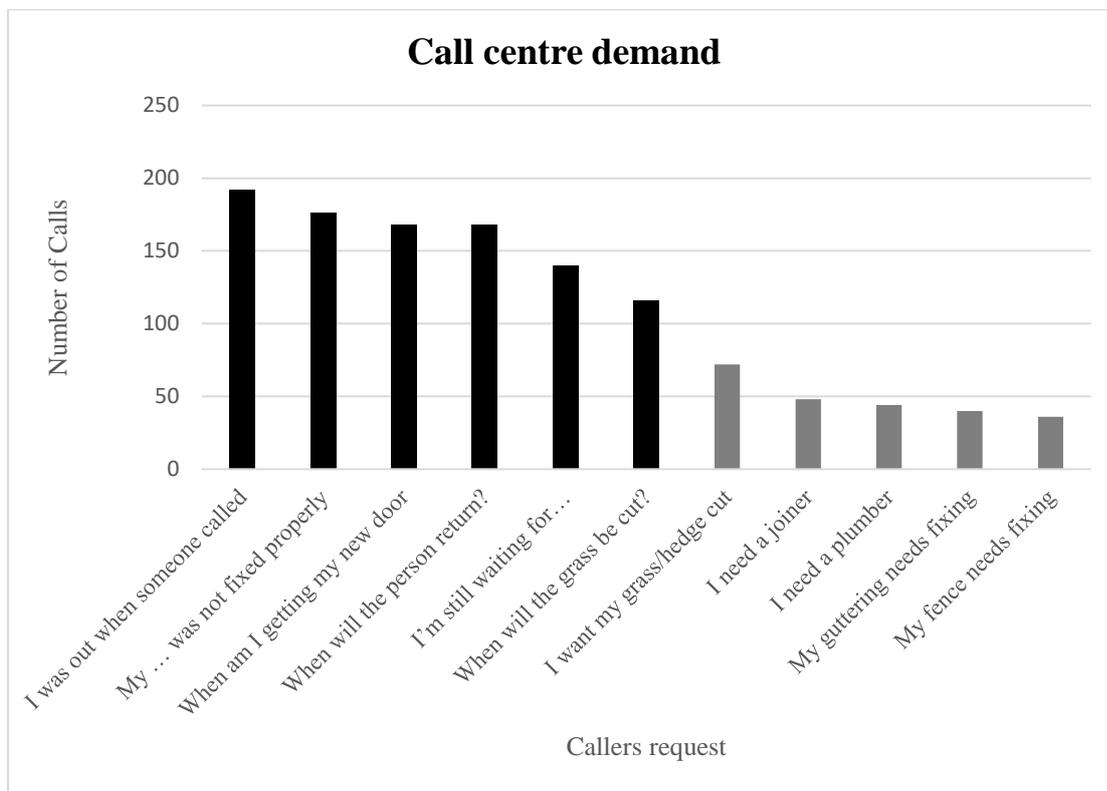
Figure 2: Calls to the 111 Service, 3 Nov. 2014 to 1 Feb. 2015 (Adapted from NHS England, 2015)
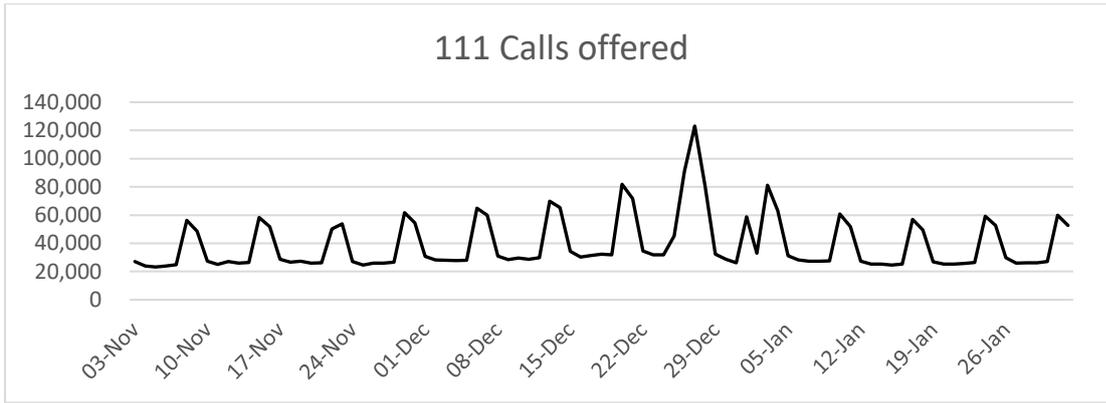
Figure 3: Abandoned calls at the 111 Service, 3 Nov. 2014 to 1 Feb. 2015 (Adapted from NHS England, 2015)
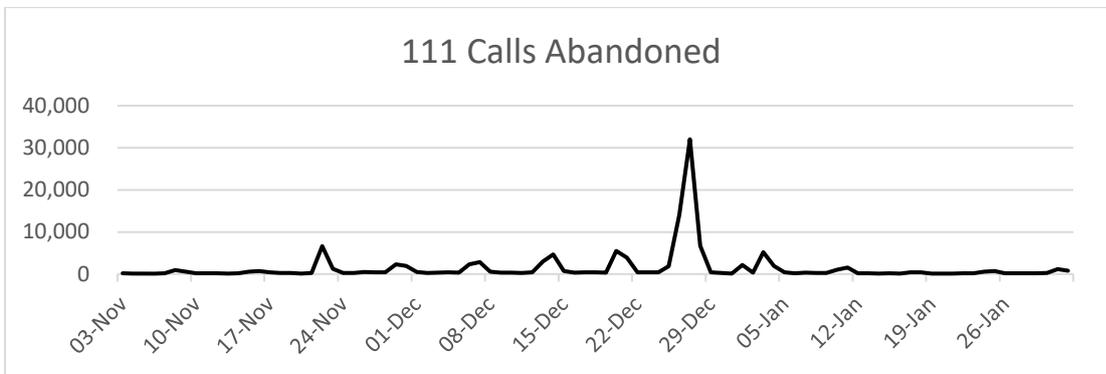
Figure 4: 111 patient referrals to A&E (daily) (Adapted from NHS England, 2015)