

Could the concept of failure demand help improve the NHS?

Prepared for The Health Foundation

April 2026

This work is licensed under CC BY-NC 4.0. To view a copy of this license, visit <https://creativecommons.org/licenses/by-nc/4.0/>

Authors

Alison Turner: conceptualisation; investigation; writing – original draft

Steven Wyatt: conceptualisation; writing – review and editing

Fraser Battye: conceptualisation; roundtable facilitation; writing – review and editing

Peter Spilsbury: supervision

With support from

Francesca Cavallaro and Ellen Coughlan, Health Foundation

Contents

Acknowledgements.....	1
Foreword	2
Executive summary	3
Key findings.....	3
Implications	4
How to read this report	5
Introduction.....	6
The productivity challenge	6
Our approach.....	7
PART 1: Understanding failure demand	8
Introducing the concept of failure demand	9
The costs of failure demand.....	11
The hidden burden on patients and carers.....	11
Workforce sustainability	12
Population health and inequities	13
System performance and opportunity costs.....	13
Why failure demand persists.....	15
PART 2: Applying the concept of failure demand	16
How failure demand manifests in healthcare.....	17
The typology in detail.....	18
Failure demand as a lens on value and productivity	21
The potential value of failure demand as a lens.....	22
Potential limitations of failure demand as a lens.....	22
Implications	24
For policymakers.....	24
For practice	24
For research.....	26
Conclusions	27
Recommendations.....	28
References.....	29

Acknowledgements

We extend our sincere thanks to all who shared their experience and insights with us. Their contributions have been invaluable in shaping our understanding of failure demand. Any errors, omissions, or misinterpretations in this report are entirely our own.

Helen Bevan	NHS Horizons
Stephen Black	Black Box Data Science
Anita Charlesworth	Health Foundation
Ron Charlton	Changing Futures Gateshead
Hamish Dibley	Hamish Dibley Consulting
Simon Dodds	SAASoft
Henry Dowell	NHS England
Nick Downham	Cressbrook Ltd
Anne Ferrey	University of Oxford
Katherine Halliday	Getting It Right First Time
Cat Harrison	Catherine Harrison Coaching and Consulting
Brin Hodgkiss	NHS England
Ibrar Hussain	Vanguard Consulting
Jacob Lant	National Voices
Alison Leary	London South Bank University
Tom Lewis	Royal Devon University Hospital
Toby Lowe	Centre for Public Impact
Brendan Martin	Public World
Clare Maxted	Gloucestershire Health and Care NHS Foundation Trust
Bradley Mellor	Mid Yorkshire Teaching NHS Trust
Matthew Mezey	Q Community
John Mortimer	Impro Consulting
Martin Myers	Getting It Right First Time
Rob Parnaby	County Durham and Darlington NHS Trust
Stephen Rocks	Health Foundation
John Seddon	Vanguard Consulting
Mark Smith	Independent consultant; Manchester Metropolitan University
Sam Spencer	Royal Borough of Kensington and Chelsea
Tanya Stacey	Gloucestershire Health and Care NHS Foundation Trust
Thea Stein	Nuffield Trust
Benjamin Taylor	Red Quadrant
Tom Underwood	NHS England; Outcomes Value & Stewardship Network
Helen Vine	Association of Ambulance Chief Executives
Paul Walley	Open University

Foreword

You know it when you see it. And, once you know the concept of failure demand, you see it everywhere. Bouncing around in a call centre queue, waiting impatiently to speak to someone who doesn't want to speak to you; among other things, you may think: 'I am failure demand'.

Sometimes you know it by its absence. A smooth transaction where you get what you were after, minus the sadly usual character-building friction. Maybe you feel the quiet surprise that everything 'just worked'.

Do we know failure demand in the NHS? What is obvious while waiting in a call centre queue becomes fuzzier when applied to healthcare services. The purpose of a call centre is obvious, and we know when it is failing to provide value. But the purpose of the NHS is a question that could consume multiple professional lifetimes.

Spotting failure demand in healthcare seems like a taller order than spotting it in the market.

And yet we feel it. Making the umpteenth call to coordinate the apparently uncoordinated care of a loved one; receiving a letter in the post for an appointment which has passed; attending a follow-up appointment but leaving uncertain as to why the appointment was needed. Are these examples of failure demand?

Perhaps this shows that failure demand has intuitive appeal and easy understanding - even if it defies a sharp-edged definition? Maybe this is where its value lies? Or perhaps it is that everyone involved can see the waste of scarce resources and human life? Is that the power of failure demand: that it forms a rare occasion when everyone would agree that there are no winners?

We entered this project with these founding intuitions and questions. Our task was to map and define failure demand, before exploring its potential value in application. Would a concept with apparent power in a market context prove useful in the very different context of the NHS?

This report is the result. It presents our findings having engaged with experts, reviewed the literature, and chewed through this material at a collaborative workshop.

It may be our result, but it is far from the final word. And one obvious area for development is improving NHS productivity – a topic currently being explored by the Health Foundation's 'Productivity Commission'.

Might failure demand offer another positive way of engaging people in this debate? A way that does not leave people with the impression that 'productivity' equals working ever harder? A way that adds value to existing concepts, such as technical or allocative efficiency?

The reader will judge for themselves. And one test they might employ is the extent to which they begin to see failure demand as they go about their lives, and in their interactions with health and care services.

If the report has done its work, its readers will be left seeing differently.

Fraser Battye, Head of Policy, Strategy Unit

Executive summary

Failure demand in healthcare: a diagnostic lens for productivity and value?

The NHS faces a persistent productivity challenge. Despite increasing resources, outputs have not kept pace with demand. Conventional responses – doing more, faster – may be missing a more fundamental question: how much of what the NHS does is self-inflicted, arising from its own poor design and processes?

Failure demand – the additional work created when services fail to meet people’s needs effectively the first time – offers a different way of understanding this challenge. First described by John Seddon in the private sector, the concept has been applied in healthcare to describe avoidable contacts, appointments and workload arising from missed opportunities, poor coordination, and system design failures. Examples include repeated contacts from communication failures; additional utilisation during extended waiting times; or people reaching crisis point before being eligible for support.

This report, commissioned from the Strategy Unit by the Health Foundation, examines whether failure demand can be more clearly defined and whether it offers a useful lens for understanding productivity and value in the NHS. Drawing on literature, expert interviews and a stakeholder roundtable, we developed a typology of failure demand signals and examined their causes, costs, and implications for policy, strategy and practice.

Key findings

The concept is intuitively resonant, but the knowledge base is fragmented.

Across the evidence and stakeholder engagement, we found strong recognition that a considerable proportion of healthcare activity may be generated by a failure to get things right the first time. However, the knowledge base is fragmented, terminologies vary, and empirical evidence is limited. The concept also lacks a clear and shared definition, meaning that different attempts to quantify failure demand are rarely measuring the same thing. As a result, contemporary use of the concept in healthcare relies heavily on interpretation and expert opinion rather than a shared theoretical foundation.

Failure demand manifests in four broad ways

We identified four types of signals: operational frustrations; rule-bound services; delays and blockages; and disjointed care. These categories reflect the diverse ways in which system design can create avoidable activity and amplify demand across organisational boundaries.

Costs are substantial and often borne by patients

Evidence suggests that the costs of failure demand could be substantial but are often hidden. These include not only the direct costs of unnecessary activity, but also the amplification of demand across services, the burden shifted onto patients and carers, and the toll on staff wellbeing and retention. Recognising this asymmetry is essential when assessing value.

While failure demand resists precise definition, this ambiguity can be a strength. Used as a diagnostic lens, it can support learning, shared sense-making and meaningful conversations about value.

Failure demand could help explain the productivity paradox

Several stakeholders suggested that failure demand could help explain why the NHS continues to consume increasing resources without commensurate improvements in outcomes. If failure demand is not simply present but growing – driven, for example, by increasing system complexity, workforce pressures, or the cumulative effects of fragmentation – then a rising share of activity may be absorbed by remedial work rather than contributing to better outcomes. While this hypothesis has not been empirically tested, it offers a plausible account of why substantial increases in resources have not translated into proportional gains. The concept challenges conventional productivity metrics by highlighting how much system activity is driven by unresolved need.

A proposed typology for shared understanding of failure demand

Our exploratory typology of failure demand signals aims to support practitioners, patients, leaders and policymakers to recognise different forms of avoidable activity and identify appropriate responses. It is designed as a starting point for purposeful improvement.

Implications

- **Policymakers** should avoid inadvertently generating failure demand through narrowly defined targets or access policies. Given the ambiguity, turning failure demand into a performance metric would be counterproductive and may lead to gaming. Instead, policymakers should create the conditions for learning, experimentation and system redesign – investing in systems thinking, enabling transparency, and aligning incentives across organisational boundaries.
- For **practitioners and system leaders**, the challenge is to identify and respond to failure demand in a way that matches its underlying causes. Some types of failure demand are essentially transactional: clear, contained problems within a single service that lend themselves well to established quality improvement methods and can often deliver tangible gains quickly. Other forms of failure demand are more complex and relational, arising from gaps between services, unclear responsibilities, or the interaction between clinical, social and behavioural needs. These require adaptive, cross-boundary approaches that emphasise learning, iteration and collaboration rather than technical fixes.
- For **researchers**, a key contribution lies in developing shared ways of conceptualising and studying failure demand that are sensitive to context and ambiguity. Rather than seeking simple measures, this work should focus on strengthening the evidence base—particularly around the burden of failure demand on patients and carers, and how system redesign might reduce avoidable activity and improve outcomes—while avoiding approaches that could be misused as targets or performance metrics.

Failure demand is not a new concept, but it may be newly relevant. As the NHS seeks to improve productivity and value, failure demand offers a timely and practical lens for understanding where effort is being wasted, and how services might be redesigned to work better for patients, staff and the system.

How to read this report

This report is structured in two parts. Part 1 (Understanding failure demand) examines what failure demand is, what it costs, and why it persists. Part 2 (Applying the concept) presents an exploratory typology of failure demand signals and explores its relationship with existing frameworks for understanding value and productivity. The report concludes with implications, recommendations, and areas for further research.

Not every reader will need every chapter. Here is a guide to what each covers:

- ***Introducing the concept of failure demand*** defines failure demand, positions our approach within the wider literature, and explains the scope of the study. Readers already familiar with the concept may wish to move straight to the next chapter.
- ***The costs of failure demand*** examines four dimensions of cost: the hidden burden on patients and carers, workforce sustainability, population health and inequities, and system performance. This is a core chapter.
- ***Why failure demand persists*** explores four structural drivers of failure demand: fragmentation, defensive pressures, misalignment of skills to needs, and mis-identification of what patients actually need. It provides important context for the recommendations.
- ***How failure demand manifests in healthcare*** presents our exploratory typology of failure demand signals, with illustrative examples. This is a core chapter.
- ***Failure demand as a lens on value and productivity*** connects failure demand to existing frameworks, including waste typologies and economic efficiency concepts. It is aimed at readers with a particular interest in the conceptual foundations.
- ***Implications*** discusses what the findings mean for policymakers, practitioners and researchers.

Readers short on time may wish to focus on the executive summary, recommendations, and conclusions. For a fuller but focused read, we suggest the introduction, the costs of failure demand, the typology chapter, and the conclusions, which together provide the core argument and practical framework. The full report, including detailed evidence and stakeholder perspectives, is for those with a deeper interest in the conceptual and methodological foundations.

Introduction

John Seddon coined the concept of failure demand (Seddon et al., 2019) to describe the additional work generated from failing to meet a customer's need sufficiently the first time. Failure demand stands in contrast with **value demand**, which relates to a service's core value. This is the work the system is designed for. Whereas failure demand is the extra burden created because the system failed to do something properly the first-time round.

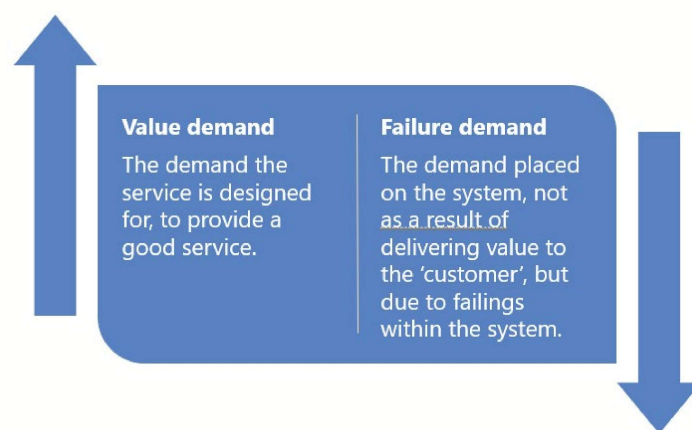


Figure 1: Value and failure demand (from Vanguard Consulting) <https://beyondcommandandcontrol.com/>

The concept was originally developed in a commercial context but has since been used in healthcare to describe avoidable contacts, appointments and workload that arise from missed opportunities to fully address patient needs. Although the concept has been in use for some time, the published literature remains surprisingly sparse. Terminologies vary, and empirical evidence is limited. As a result, contemporary use of the concept in healthcare relies heavily on interpretation and expert opinion rather than a shared theoretical foundation.

Failure demand derives from systemic issues baked into the design of processes, services and interfaces which hinder attempts to manage demand and which warrant deeper exploration (Walley et al., 2019). Failure demand – through missed opportunities, delays, and unmet needs – is likely to lead to productivity losses. But can failure demand serve as a meaningful framework for understanding productivity and value in healthcare?

The productivity challenge

As failure demand represents avoidable work, it is inextricably linked to productivity. Policymakers view boosting productivity as essential to sustaining the NHS and supporting the wider economy (Moody and Powell, 2025). The Government's [10 Year Health Plan for England](#) commits to a 2% year-on-year productivity gain over three years (Moody and Powell, 2025).

Improving efficiency and reducing waste is also a priority for the public, particularly improved coordination between services (Benniche et al., 2025). The National Audit Office recommends public services explore services from a user perspective to get services right first time and reduce

failure demand (National Audit Office, 2025). Initiatives such as Getting It Right First Time (GIRFT) and Making Every Contact Count can be seen as addressing manifestations of failure demand, though they are not always explicitly framed as such.

Whilst NHS productivity grew between 2021/22 and 2022/23, the gap compared to pre-pandemic levels is significant (Arabadzhyan et al., 2025). The challenge to meet the government's 2% productivity target is substantial, as illustrated by expert elicitation for the Health Foundation's [NHS Productivity Commission](#) (Cook et al., 2025; Fozzard et al., 2025). Experts gave just a 15% chance of success in the short term. Their central estimate: 0.7% annual growth over five years - is well below the 2% target. Even over 6-10 years, estimates only reached 1.4% per year.

The NHS faces systemic barriers to achieving productivity gains. Inequalities persist, with poorer areas seeing lower healthy life expectancy (Health Foundation, 2023, updated 2025). Waiting lists remain high despite modest improvements (Opie-Martin et al., 2025). Public satisfaction remains low - only 21% report being 'very' or 'quite' satisfied with the way in which the NHS runs (Taylor et al., 2025) - and over 460,000 staff report experiencing burnout daily (West and Wallbank, 2025).

Our approach

Funded by the Health Foundation, we explored whether failure demand provides a meaningful lens through which to understand opportunities to improve efficiency, productivity and value in the NHS. Our approach was iterative, integrating experiential and empirical evidence with stakeholder engagement:

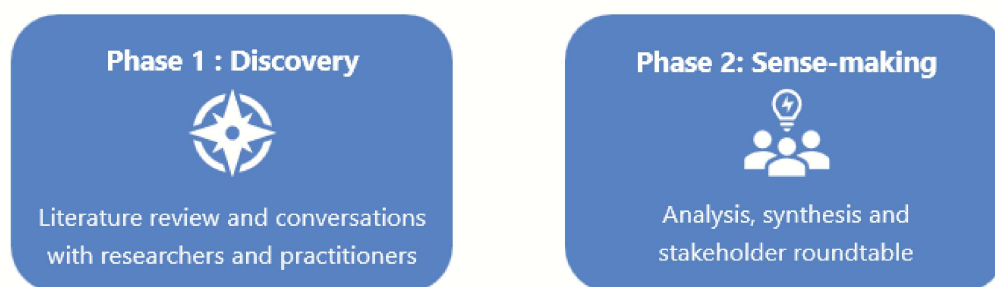


Figure 2: Our approach

In phase one, we explored the concept of failure demand through a review of published literature and informal interviews with researchers and practitioners. In phase two, we synthesised findings and refined our thinking via a roundtable with clinicians, improvement specialists, academics and system leaders.

This report approaches failure demand from a system and policy perspective, offering a structured typology of its signals, exploring why it persists, and examining its costs. We set out recommendations for policymakers, system leaders and researchers, and intend for the report to

provide a springboard for debate and a foundation for further analysis and research at system, organisation and service level.

PART 1: Understanding failure demand

Introducing the concept of failure demand

Failure demand in healthcare is typically described as the additional contacts, appointments and activities that arise when underlying patient needs are not fully met the first time. As Walley et al. (2019) explain, *"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"* - a framing that highlights how inefficiencies and misdirected effort can compound rather than resolve patient need.

In the words of John Seddon, failure demand is *"a concept that is easily understood and easily misunderstood"* (Seddon et al., 2019). One way to clarify it is by distinguishing it from related forms of demand. Randle and Kippin's typology offers a helpful framework (Randle and Kippin, 2023).

Failure	Demand caused by errors or poor processes
Avoidable	Demand arising from behaviours that can be changed
Excess	Demand created by providing more than is needed
Preventable	Demand arising from causes that could be removed earlier
Co-dependent	Demand that is unintentionally reinforced by dependence

Table 1: Typology of demand (Randle and Kippin, 2023)

In practice, the concept has been interpreted at very different scales. Some adopt a "zoomed-in" lens, using quality improvement tools to address specific process failures. Others take a "zoomed-out" view, linking failure demand to broader questions of purpose, complexity, and inter-agency dynamics.

A more expansive argument (Hensher et al., 2020) suggests that a substantial share of healthcare demand is itself a downstream consequence of failures in the wider economy - preventable ill-health arising from smoking, alcohol, obesity and pollution. From this vantage point, a patient presenting with a diet-related chronic condition is experiencing failure demand because their need reflects failures to prevent avoidable harm earlier in the system. This broader framing situates healthcare activity within its wider determinants, but it may be too expansive to be actionable. Any workable definition of failure demand needs to address the question of prevention carefully, and we would argue should err on the side of a narrower scope - otherwise the concept risks expanding to encompass virtually all healthcare demand, at which point it loses its analytical value.

In this paper, we adopt a narrower and more operational focus: failure demand generated within or between health and care services, where demand is generated because the patient's need was not met effectively the first time. This tighter definition is more tractable for service leaders, commissioners and improvement teams, while still reflecting the broader insight that not all demand placed on healthcare is either legitimate or inevitable.

While perspectives varied across our interviews and roundtable, a common thread was the recognition of failure demand as a signal - not just of inefficiency, but of deeper design flaws. Whether focused on flow, relationships, or system design, most agreed that failure demand is best understood as part of a wider conversation about value. The concept has a natural affinity with value-based healthcare, which defines value as outcomes that matter to patients relative to the cost of achieving them: failure demand represents activity that increases cost without improving outcomes, eroding value on both sides of the equation. We explore this relationship further in the section on [failure demand as a lens on value and productivity](#).

The Challenge - and Value - of Ambiguity

Across the literature and our empirical work, a recurring theme is the ambiguity at the heart of the concept. Failure demand is widely recognised in practice, yet consistently difficult to define with precision. Questions about its boundaries quickly surface: could all preventable diseases be defined as failure demand? Is over-testing failure demand, or a reflection of shifting societal risk appetites? If the concept is defined too broadly, it risks losing its analytical utility.

The concept's power may lie in its ability to provoke useful conversation, rather than in achieving a watertight definition. As one roundtable contributor put it, it is a "tantalising concept" - intuitively recognised but resistant to neat categorisation. Failure demand could function most usefully as a lens to prompt inquiry, reframe assumptions and stimulate reflection on how systems meet - or fail to meet - people's needs.

Concerns also emerge about the term itself. The word "failure" can be counterproductive: frontline staff may feel blamed, and organisations may resist transparency. Alternative framings such as "getting it right first time" or value-focused terminology may provide a more constructive starting point. Any future application would benefit from careful framing, or possibly rebranding, of the term.

The costs of failure demand

The costs of failure demand extend across at least four dimensions: the invisible burden shifted onto patients and carers when services fail to resolve needs; the impact on workforce wellbeing and turnover; ill-health and inequities that could potentially have been avoided; and the direct costs of unnecessary activity and indirect costs of demand amplification across services and over time.

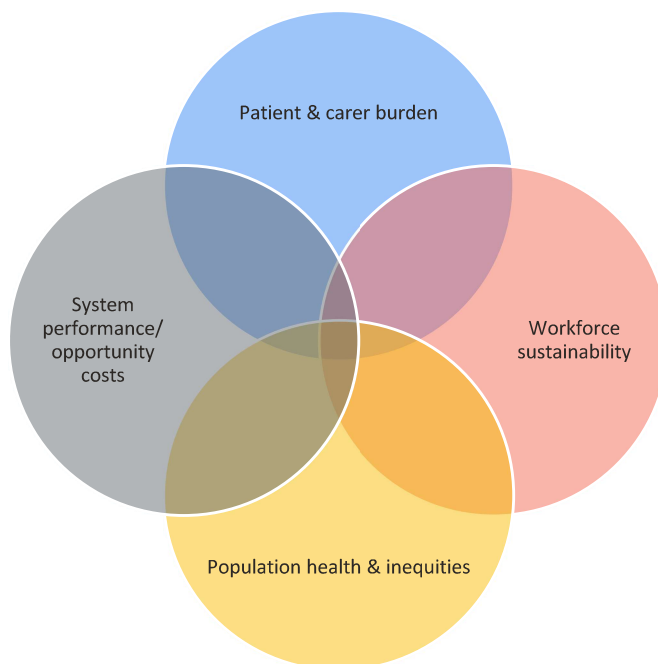


Figure 3: Costs and consequences of failure demand

These four dimensions are not independent. They interact and compound one another in ways that are largely invisible to conventional measurement. The hidden burden on patients and carers masks demand that would otherwise be visible to services; workforce attrition driven by purposeless work reduces system capacity to address needs effectively, generating further failure demand; and health inequities created by standardised service models produce concentrated patterns of crisis demand that amplify pressure across the system. The direct costs captured under system performance are therefore only the most visible layer of a deeper structure of loss.

The hidden burden on patients and carers

One of the strongest themes from the roundtable was the unmeasured cost that failure demand imposes on patients and carers. Current NHS metrics capture activity and waiting times, but they do not capture the burden transferred to individuals: the time spent chasing referrals and repeat appointments, the financial costs of travel and lost earnings, the emotional toll of navigating fragmented services, and the opportunity cost of energy diverted from work, family, and recovery into managing the system itself.

Participants described patients “going into battle” with the system, and carers giving up work to navigate services on behalf of relatives. One example given was end-of-life care, where much of the navigation cost is shifted to carers and families. Other examples included patients being asked to recall their own medication histories because systems do not share data, and the emotional and practical toll of repeated failed contacts with services.

This has a paradoxical implication for productivity measurement. Because the burden on patients is invisible to NHS metrics, any improvement work that reduces this burden (for example, improving the NHS App so patients no longer need to track their own records) could register as a cost to the system without a corresponding measurable benefit. Current activity metrics treat every appointment, test, and referral as productive output, regardless of whether it arose from genuine need or from a previous failure to meet that need. As one workshop participant put it, improvements in patient-facing failure demand will “just look like a cost rather than a productivity gain” because the current measurement framework does not account for burden borne outside the system.

Workforce sustainability

Several interviewees referenced the potential impact on staff wellbeing, turnover and retention. When staff are repeatedly required to respond to avoidable demand, it can erode their sense of purpose and professional identity. Some clinicians report feeling reduced to mechanised roles, following rigid protocols rather than exercising their judgement or delivering holistic care (Montori, 2020; Shah and Foell, 2024). One practitioner described an occupational therapist who broke down in tears upon realising how far her day-to-day tasks - limited to installing ramps - had drifted from the values and training that drew her to the profession. This disconnect between professional ethos and operational reality can contribute to burnout and attrition.

Staff might feel a persistent sense of complicity in a system that fails to meet people’s needs effectively. In high-failure-demand environments, defensive practice can become the norm (Downham et al., 2025; Beardsell, 2025): clinicians may rely on forms and checklists not as tools for care, but as shields against risk. While these mechanisms offer a sense of protection, they can also stifle meaningful engagement with patients and further alienate staff from their core purpose.

Conversely, addressing failure demand can unlock improved wellbeing. Several practitioners reported that staff regain a sense of professional agency when empowered to explore failure demand. However, not all staff feel able to challenge failure demand. Moreover, leaders who attempt improvement sometimes do so at personal cost, taking on the burden of systemic change without adequate support. For failure demand to be meaningfully addressed, cultural and structural shifts are needed to ensure that those driving change are not left carrying the weight alone.

Population health and inequities

Health inequities frequently came up in conversation with practitioners, specifically the distinction between equality and equity. Equality-based approaches provide the same offer to all patients; equity-based approaches proportion support to need. For much of the population, an equality-based model functions adequately. But for patients with multiple or complex needs – those whose conditions interact, whose circumstances compound their clinical needs, or who face barriers to navigating standard pathways – a standardised offer generates failure demand by design.

Spyropoulos et al. (2022), evaluating a multi-agency partnership supporting people experiencing multiple disadvantage in Stoke-on-Trent, found that services calibrated for the average user systematically excluded those with the most complex needs not through deliberate gatekeeping, but through the cumulative effect of processes, thresholds and eligibility criteria that were not designed with this population in mind. The result was a revolving door of crisis presentations and re-referrals: a pattern recognisable as failure demand operating across organisational boundaries. Whilst this study was not specific to healthcare, there is an obvious parallel with services for people with multiple long-term conditions, learning disabilities, or co-occurring mental and physical health needs.

System performance and opportunity costs

Failure demand has significant implications for system performance, often creating what might be described as a demand amplification trap. Each of these additional contacts registers as activity in national datasets, inflating headline figures while obscuring the underlying issues that generated them. Failure demand is rarely contained. When services are unable to meet needs effectively the first time, those needs don't disappear, they resurface, often in more urgent, complex, or costly forms.

Appointment volumes in general practice are at record highs, yet patient satisfaction with access continues to decline (Sinnott et al., 2025). A GP-led analysis of 599 consultations in a large practice found that 19% arose from what the authors described as system failures, stemming from ineffective models of care and processes. While the generalisability of this finding is uncertain, it is consistent with a broader pattern recognised by practitioners: that a meaningful proportion of general practice workload may be generated by unresolved or poorly coordinated care rather than new clinical need. There is growing evidence that continuity of care improves productivity in general practice, particularly patients who are older, have multiple long-term conditions or mental health conditions (Kajaria-Montag et al., 2023; Sinnott et al., 2026).

The ripple effects extend across the system. A matched cohort study by James et al. (2024) found that patients waiting over 18 weeks for elective treatment generated significantly more healthcare contacts than matched controls, particularly in general practice and secondary care. In one

example, patients on a respiratory waiting list generated nearly 8,000 additional secondary care contacts in a year, costs that were not necessarily borne by the service responsible for the delay. The authors describe this as a “false economy”, where deferred treatment appears to save money but ultimately drives higher system-wide costs. Recent linked data analysis from the Health Foundation's Networked Data Lab (Prendergast et al., 2025) reinforces this picture: people on elective waiting lists used primary care at a higher rate than the general population, with the weekly cost of GP appointments for those waiting substantially exceeding the cost of A&E attendances. Patients who waited longer also saw relatively greater increases in primary care and emergency care use after treatment, suggesting that prolonged waits may generate additional downstream demand that persists beyond the waiting period itself.

This dynamic is even more starkly illustrated in Changing Futures' work (Charlton, 2025), which traced one individual's 14-year journey through multiple public services, accruing nearly £2 million in costs through repeated crisis presentations. Following a shift to intensive, relationship-based support, annual costs dropped to £13,000. While this case reflects an extreme, the underlying pattern - unresolved need leading to escalating, cross-system demand - is familiar across NHS settings. Crucially, these costs are rarely visible to the services that generated them, making the true impact of failure demand difficult to quantify and easy to overlook.

Why failure demand persists

Failure demand is described in the literature as an unintended consequence of how work is designed (Jones and Pereira, 2024; Sinnott et al., 2025). Understanding the contributing factors is critical to tackling and preventing failure demand. To support this, we draw on a framework proposed by Downham (Anderson-Wallace and Downham, 2024) whose work on failure demand in healthcare is grounded in professional experience. His framework identifies four sources of failure demand: fragmentation by design; defensive pressures; specialist, generalist and citizen muddle; mis-identification of needs. We found these categories useful because they resonated with themes from our own interviews and roundtable and align with the wider literature on system design. We use them as an analytical scaffold rather than an empirically validated taxonomy.

Fragmentation by design was the most frequently cited root cause across our interviews and roundtable. Siloed budgets, vertical accountability structures, and IT systems that cannot share data between organisations mean that patients frequently fall between services. Roundtable participants offered vivid examples: duplicate scans carried out because results cannot be shared between providers; palliative care patients whose medicines cannot be supplied out of hours because no single service owns the problem; and patients with co-occurring conditions (such as autistic spectrum disorder and addiction) who fall below the threshold for every specialist team and cannot access a coherent pathway.

Defensive pressures - the tendency to manage risk through avoidance rather than judgement - were also highlighted. One roundtable participant noted that the NHS is “good at avoiding risk rather than managing risk”. While defensive over-testing and over-referral may initially appear closer to excess demand than failure demand, the two are closely linked in practice. Unnecessary investigations consume capacity, contribute to longer waits, and can trigger cascades of further testing and referral, each of which generates downstream demand that would not otherwise have existed. Defensive practice may therefore be better understood as a mechanism that produces failure demand rather than as failure demand itself.

The **specialist, generalist and citizen muddle** describes the misalignment of skills to needs. One example concerned ventilated patients admitted via ambulance for routine catheter changes manageable in the community - a predictable need met by an emergency response. This prompted one participant to propose a complementary concept of “failure supply”: the system’s failure to organise the right resources to meet known demand. We also heard from experts about the tendency in the NHS to place less experienced practitioners at the front door to triage patients (the “rookie factor” (Leary, 2024)) which can slow down decision-making and patient flow.

Mis-identification of needs arises when services define demand in their own terms rather than understanding what the patient actually needs. Roundtable participants stressed the importance of starting from the question “what matters to you?” and understanding demand from the user’s perspective.

PART 2: Applying the concept of failure demand

How failure demand manifests in healthcare

Given the lack of a shared understanding of failure demand in healthcare, we aimed to clarify how and where it manifests in practice. Drawing on literature and expert interviews, we identified a wide range of examples — from isolated process failures to systemic issues in care pathways. To make sense of these, we developed a typology outlining the different forms failure demand can take, which was further refined through feedback at a stakeholder roundtable.

To bring these different examples together in a way that supports clearer thinking and practical discussion, we organised them into a simple matrix. The matrix is built around two dimensions. The first (the x-axis) distinguishes between **transactional interactions**, where someone is seeking help with a specific issue ("fix me"), and **relational interactions**, where needs are more complex, ongoing or shaped by wider circumstances ("help me"). The second (the y-axis) distinguishes between failure demand that arises **within a single service** and failure demand that emerges from the interaction **between services** and gaps between organisational responsibilities. Although we name the extreme point of these axes ("fix-me"-"help-me" and within-service to between services), we do not see these as dichotomies, but as graduated scales.

Like any model, the typology simplifies reality. Its value lies in offering a shared framework to help teams interpret what they're seeing and structure conversations about the causes and responses to different types of failure demand, rather than serving as a rigid or definitive classification.

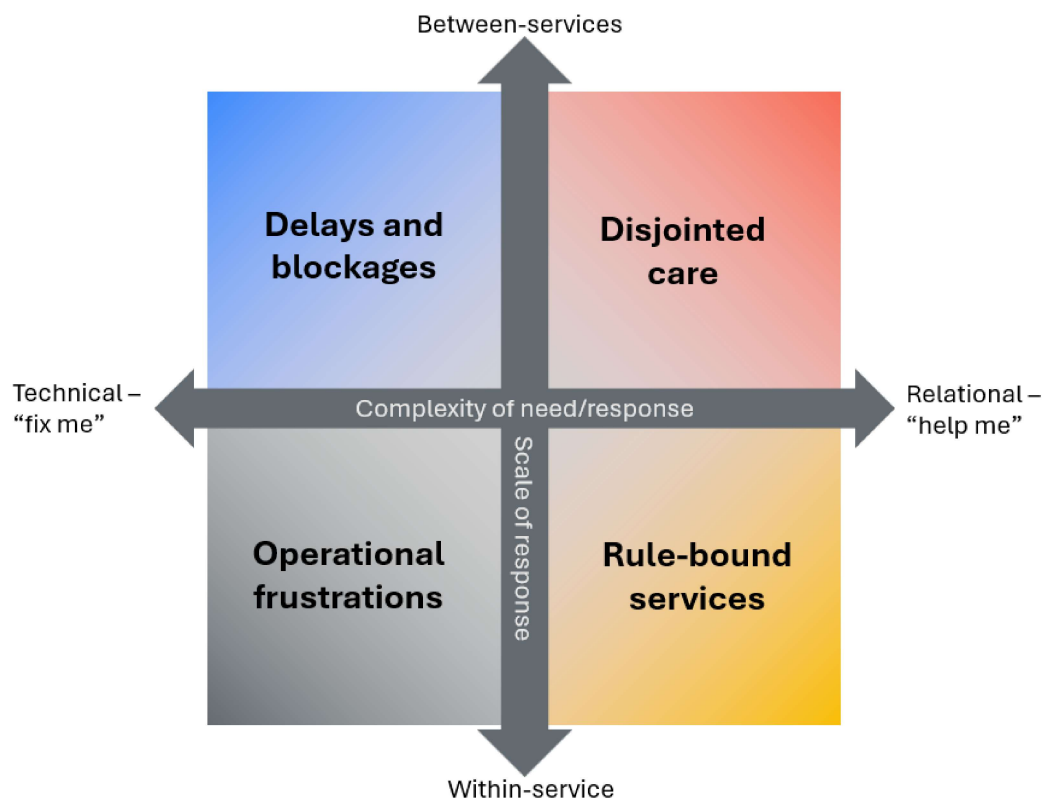


Figure 4: Signals of failure demand: a typology

The typology in detail

Operational frustrations			WITHIN-SERVICE
SUB-CATEGORY	DESCRIPTION	ILLUSTRATIVE EXAMPLE	
Booking breakdowns	Multiple attempts to book an appointment due to system restrictions or poor design.	NHS inflexibility is behind falling vaccine uptake (Wilkinson, 2025)	
Communication gaps	Incomplete or unclear information that requires patients to make repeated contacts for clarification.	Lost in the system: the need for better admin (Cream et al., 2025)	
Supply issues	Poor availability of prescriptions/equipment, requiring multiple contacts to resolve.	Your experiences of getting prescription medication (Healthwatch, 2023)	
Rule-bound services			WITHIN-SERVICE
SUB-CATEGORY	DESCRIPTION	ILLUSTRATIVE EXAMPLE	
Eligibility thresholds	Unclear or overly rigid eligibility criteria resulting in inappropriate referrals or patients not receiving care until crisis point.	Falling through the gaps? A closer look at NHS continuing health care. (Hutchings et al., 2025); Evaluation of the Changing Futures programme - baseline report. (CFE Research and Cordis Bright, 2023)	
Support gaps	Essential support - e.g. help with independent living, interpretation, lifestyle change - is absent by default, leaving gaps that become visible when individuals reach crisis or re-present.	The State of Health and Care of Older People in England 2024. (Reeves et al., 2024)	

Delays and blockages			BETWEEN-SERVICES
SUB-CATEGORY	DESCRIPTION	ILLUSTRATIVE EXAMPLE	
Flow and duplication	Breaks in patient flow create bottlenecks and avoidable duplication - e.g. repeat tests to confirm surgical suitability after long waits.	Hospital of the future: ending the patient gridlock. (Beacon and Seamark, 2025)	
Long waits	Additional service utilisation or reduced efficacy of treatment resulting from extended waits - including re-referrals and deterioration during the wait.	The cost of keeping patients waiting: retrospective treatment-control study of additional healthcare utilisation for UK patients awaiting elective treatment. (James et al., 2024)	
Disjointed care			BETWEEN-SERVICES
SUB-CATEGORY	DESCRIPTION	ILLUSTRATIVE EXAMPLE	
Poor transitions	Gaps at service interfaces - e.g. child-to-adult services or hospital-to-community - create unmet needs.	Experiences of transition from children's to adult's healthcare services for young people with a neurodevelopmental condition. (Shanahan et al., 2021)	
Loss of continuity	Fragmentation obscures repeat contacts, meaning missed opportunities to diagnose or treat.	Delayed diagnosis of cancer: a thematic review of general practice indemnity (Thomas, 2025)	
Pixelation	Loss of the nuanced, whole-person picture as individuals move through multiple services, concealing complex needs.	The Public Services Burning Platform: Using Existing Transactional Data to Make a Case for Relational Public Service Approaches. (Charlton, 2025)	

Table 2: Signals of failure demand: detailed typology

Understanding failure demand is easier when we can see it. The following vignettes illustrate how different signals in the typology emerge in practice — from missed opportunities to help people the first time, to delays, duplication and unclear responsibilities. Each example shows how system design can generate additional work and frustration for staff, patients and families.

Delays and blockages

A newly admitted patient encounters small, repeated delays - diagnostics that only run intermittently, specialist reviews that must be repeated because the right clinician wasn't available earlier, and multiple referral hand-offs that restart when shifts change - that accumulate into days spent in hospital despite the patient being medically fit. Bottlenecks at key transition points - between A&E and the wards, between teams inside the hospital, and between hospital and community services - create blockages that ripple backwards, slowing flow and generating failure demand (Beacon and Seamark, 2025).

Operational frustrations

A new mother tried to book her baby's routine 8-week vaccination but found herself caught in an inflexible system. She was repeatedly told to "call back later" because appointment slots were not yet open, or to try again on specific days. When she called at the instructed times, the slots had already filled. Despite receiving automated reminders urging her to book, she made four unsuccessful attempts before resorting to a written complaint to secure an appointment (Wilkinson, 2025). This experience reflects how rigid booking processes, limited appointment availability, and poor coordination can generate failure demand.

Disjointed care

Parents can often end up taking on the role of care navigator as their children transition to adult services. Poor communication often means gaps in care - a recent study provides an example of an 18-year-old with a neurodevelopmental disorder. Despite his mother trying on numerous occasions to contact the adult ADHD services, no information was provided until about 5 months after the young person's 18th birthday. In the meantime, they had felt obliged to seek support privately, at a cost of £400 for an initial consultation and £200 per follow up appointment (Shanahan et al., 2021).

Rule-bound services

Families seeking NHS Continuing Healthcare often find themselves navigating a process that is complex, confusing and inconsistent. Many people undergo lengthy assessments only to discover they are ineligible, despite having significant ongoing needs (Hutchings et al., 2025). This uncertainty forces families, carers and clinicians into repeated contacts, appeals, reassessments and referrals as they try to understand decisions and secure the right support. The result is substantial failure demand: avoidable workload generated by unclear rules and inconsistent interpretations.

Failure demand as a lens on value and productivity

A growing body of typologies and frameworks has helped shape our understanding of inefficiency and low-value care in healthcare. But how does the concept of failure demand complement or challenge these existing perspectives?

Estimates suggest that up to 30% of healthcare activity is unnecessary or of low value (Braithwaite et al., 2020). Berwick and Hackbarth's influential framework identifies six categories of waste: failures in care delivery, failures in care coordination, overtreatment, administrative complexity, pricing failures, and fraud (Berwick and Hackbarth, 2012). While broad in scope, the first two categories closely align with the mechanisms that generate failure demand, highlighting how systemic inefficiencies can drive avoidable activity. The OECD's model (OECD, 2017) adds another layer, mapping waste to system-level drivers and actors. It distinguishes between wasteful clinical care, operational waste, and governance-related waste. Failure demand intersects with all three but is most strongly connected to wasteful clinical care and operational inefficiencies that lead to repeated or unnecessary contacts. Together, the Berwick/Hackbarth and OECD frameworks provide supply-side categorisations of waste, whereas failure demand offers a complementary demand-led lens.

Health economists have long examined unwarranted variation and inefficiency, developing related concepts such as technical and allocative efficiency, overuse, underuse and low-value care. These approaches address the same underlying issue of avoidable demand but from different angles. Technical efficiency focuses on delivering care with minimal resources (Palmer and Torgerson, 1999) yet by its nature is indifferent to whether the activity should occur at all - a system can appear highly efficient while generating large volumes of failure demand. Allocative efficiency considers whether resources are directed to their most valuable use (Palmer and Torgerson, 1999), highlighting how failure demand diverts capacity away from genuine need and preventive work. Failure demand could be described as a form of allocative inefficiency: resources are being diverted to remedial work rather than the most productive use. This reveals an important limitation of volume-based productivity measures which can mask failure demand, as this remedial activity still counts as output. Understanding the composition of demand, not just the efficiency of supply, is therefore essential — and failure demand could provide a practical framework for doing so.

Verkerk et al. developed a typology of low-value care - ineffective, inefficient, and unwanted - which offers appropriate responses - "limit" through policy, "lean" through redesign and "listen" through shared decision-making respectively (Verkerk et al., 2018). Finally, the concepts of overuse and underuse represent another lens for understanding waste and value. Hofmann frames their coexistence as a paradox that undermines system sustainability (Hofmann, 2025) while Korenstein et al. map harms caused by overuse that can trigger cascades of downstream interventions, similar to failure demand (Korenstein et al., 2018).

The potential value of failure demand as a lens

Failure demand complements existing frameworks by drawing attention to problems designed into systems themselves. At its core is a focus on demand analysis: understanding why people seek help, from their own perspective, and distinguishing between genuine need and activity generated by system shortcomings. This challenges the assumption that rising activity necessarily reflects rising demand, instead prompting a more critical examination of whether services are addressing need or simply reworking previous failures. Several roundtable participants suggested that failure demand could help explain why the NHS continues to absorb increasing resources without corresponding improvements in outcomes. In this way, it challenges conventional notions of productivity and efficiency, offering a reframing that prioritises value over volume.

The concept may resonate across professional and public audiences, offering a shared language to describe inefficiencies that often go unnamed. By highlighting how poor design - in policies, pathways, or processes - can generate avoidable work, failure demand encourages a shift in mindset: from asking how to do more, to questioning how much of what we do is even necessary.

Potential limitations of failure demand as a lens

While failure demand potentially offers a compelling lens on inefficiency, it is not without limitations. A sceptic might argue that the concept repackages existing well-understood ideas such as waste, rework and low-value care; whilst the concept may facilitate conversation, it does not offer a genuinely new understanding. A key challenge lies in its definition: a failure to deliver value at first contact. In healthcare, value is not always straightforward. From a patient's perspective, value is shaped by expectations as well as clinical need and the two do not always align. This raises a fundamental question: whose definition of value should guide our understanding of failure demand, and how do we distinguish unmet need from unmet expectation?

Identifying and quantifying failure demand remains difficult. Because its boundaries rely on judgements about value - whether a contact was genuinely necessary, whether a need could have been met earlier, whether the system or the patient bears responsibility for a repeat presentation - the concept does not lend itself easily to measurement through routine data. These are interpretive questions that depend on context, clinical judgement, and the patient's own perspective, and they are often better suited to close observation or qualitative enquiry than to routine measurement. Existing datasets rarely capture the full picture - particularly the burden absorbed by patients, carers, and other parts of the system that never surface in administrative records. As a result, organisations may only perceive fragments of failure demand, making it hard to track trends, evaluate interventions, or understand where in a pathway the load is generated.

Language also matters. Several roundtable participants noted that the term "failure" may be counterproductive in a system where staff already feel under pressure and organisations are

cautious about acknowledging shortcomings. Framing the concept in terms of value - or using more constructive language such as "getting it right first time" - may support wider engagement, with the technical terminology reserved for analytical contexts.

Another limitation lies in the assumption that earlier or better intervention could have prevented subsequent demand. In practice, the opportunity to act is not always visible at the time. Hindsight can make missed opportunities appear obvious, when in reality, clinicians may have lacked the information needed to intervene effectively. Conversely, acting too early on incomplete evidence risks overtreatment and there is little evidence to suggest that current approaches to risk prediction reliably translate into improved outcomes (Oddy et al., 2024). Failure demand should not be interpreted as implying that all adverse outcomes are preventable - clinical uncertainty is an inherent feature of healthcare. Distinguishing failure demand from legitimate iterative care is a matter of judgement, not clean measurement.

Finally, structural barriers may constrain efforts to reduce failure demand. As Hensher et al. (2020) observe, "one person's 'waste' (or failure demand) is always another's income." Reducing unnecessary activity in one part of the system may threaten the financial viability or workload of another, particularly under current payment and contracting models. Roundtable participants echoed this, highlighting how siloed budgets and cost-shifting often lead services to prioritise their own pressures over system-wide value. Considerations such as organisational capacity and culture; leadership turnover; and skills gaps will also factor. Addressing failure demand, therefore, requires more than clinical or operational change: it may require a realignment of incentives, governance, and accountability across the system.

Implications

For policymakers

Every new policy, pathway, target or digital service carries the potential to create additional demand not by meeting new needs, but by failing to resolve existing ones (Jones and Pereira, 2024; Sinnott et al., 2025). Access policy is a case in point. In general practice, a sustained focus on appointment volume and same-day access appears to have fragmented care for people with long-term conditions, weakening the continuity and relational knowledge that help prevent needs from escalating (Downham et al., 2025). The new GP contract, which introduces performance management of practice responses to urgent cases, carries a related risk: as the Health Foundation has noted {Pettigrew, 2026 #1085}, similar targets under previous contracts created perverse incentives that prioritised speed of response over appropriateness of care. A systematic review spanning 40 years of access interventions found that most targeted “permeability” - the ease of getting through the door - while neglecting relational aspects of care that influence how needs are understood and met (Sinnott et al., 2025).

Structural changes can produce similar effects. When new roles and services introduce new steps or fragment responsibility, avoidable contacts can multiply. Sinnott et al. (2026) describe cases where patients who previously accessed services (such as ear syringing and podiatry) directly were required to obtain GP referrals, creating two transactions where previously one sufficed. Digital-first models demonstrate a related risk: emerging evidence suggests that digital systems can inhibit effective communication, requiring patients to construct “digital facsimiles” that fail to convey the full breadth of need (Dakin et al., 2024; Sinnott et al., 2026).

A consistent message across the evidence is that failure demand should not be treated as a performance metric. Attempting to measure and compare risks gaming, suppressing transparency and undermining the learning potential of the concept. Failure demand should remain a diagnostic lens - a tool for inquiry, not performance management.

Policymakers nonetheless have an important enabling role. This means creating the conditions for learning, experimentation and improvement, facilitating rather than directing. As Kippin et al. (2015) note, *“we can’t manage what we don’t understand.”* Understanding failure demand requires active engagement with patients, carers and staff, combined with demand analysis using multiple methods including explicative analytics, qualitative enquiry, ethnographic observation, and patient narratives.

For practice

Failure demand challenges conventional assumptions about productivity and efficiency. Rather than asking how to do more with less, it invites a more fundamental question: how much of what we do should we not be doing at all? In theory, if a significant share of current workload is generated by system defects and delays, then designing out those failures could release capacity

for work that genuinely adds value. The argument therefore is not about squeezing more from stretched teams, but about reducing the wrong work to make space for the right work.

Emerging evidence from relational public service pilots supports this hypothesis. Initiatives focused on trust, continuity and personalised support rather than transactional, target-driven delivery have shown reductions in service use among people with complex needs (Lowe and Hesselgreaves, 2024; Fox et al., 2025; Dent et al., 2026). While some headline figures are dramatic, such as the Gateshead case where service interactions dropped from 3,300 to 20 over two years, these often involve small cohorts of specific patient populations and there is some uncertainty how well these may scale. Larger-scale evaluations, such as Community Led Support across 35 English sites, show more modest but consistent reductions in demand over time. Though not based on randomised trials, findings suggest that relational, context-aware approaches can reduce avoidable activity and improve outcomes.

A recurring question is whether failure demand can be measured. At present, the NHS lacks the tools to quantify it reliably at system level. Routine data capture activity, but not the reasons behind it nor whether needs were met. Rather than seeking a single metric, services should focus on identifying signals of failure demand through a mix of methods: demand analysis, direct observation, patient stories, and staff insight. The practical question is not “how much failure demand do we have?” but “where is it showing up, and what can we learn from it?” Identifying failure demand depends on normative learning and systems thinking to challenge long-held assumptions. Without this, there’s a risk of simply shifting the problem downstream rather than resolving it.

The distinction between signals of failure demand has a practical consequence for how teams should go about identifying and responding to it. Failure demand that arises within a single service - particularly the transactional “fix me” problems in the lower-left quadrant of the typology - are driven by relatively clear, traceable causes. For example, booking errors, stock issues or steps in a process that routinely generate avoidable rework. These problems are amenable to an analytical approach: study the data, map the process, understand where defects occur and design targeted solutions. Established quality improvement methods such as Lean and Six Sigma (Radnor and Williams, 2025; McDermott et al., 2022) are well suited to this domain.

Failure demand that arises across service boundaries - particularly the relational “help me” problems in the upper-right quadrant - span clinical, social and behavioural dimensions. The causes are emergent and entangled: they evolve from the interaction between services, between clinical and social needs, between patient expectations and system design. Here, progress depends on observation, systems thinking, learning and experimentation. Approaches such as the relational principles described by Lowe and Hesselgreaves (2024) and the intensive, relationship-based support documented by Charlton (2025) operate in this way, working iteratively with individuals rather than imposing a standardised pathway.

For research

National extrapolations from pilot data rest on untested assumptions about scalability and population comparability, and none of the evidence cited comes from randomised or quasi-experimental designs with matched comparison groups. The case for further investigation is strong, particularly given the parallels between the populations described in these pilots and NHS patients with multiple long-term conditions or co-occurring mental and physical health needs.

Given the burden of failure demand on patients and carers, their perspectives are missing from the existing knowledge base. Patients are the only people who experience the whole system end-to-end. They are uniquely placed to identify where failure demand arises, how it cascades, and what burden it imposes.

Conclusions

This project set out to test whether failure demand can be defined more clearly and whether it offers a useful lens for understanding productivity and value in healthcare. Our conclusion is that it can - with important caveats.

The concept is intuitively resonant. Across the literature, expert interviews, and the stakeholder roundtable, we found consistent recognition that a significant amount of healthcare activity may be generated not by patient need but by the system's failure to get things right the first time. Failure demand provides a unifying lens through which clinicians, patients and managers can better understand patterns of demand and identify where change is needed. For the concept to have real value, it requires acknowledging that we should be as attentive to the costs borne by patients as to those borne by the service.

Failure demand may not always incur significant costs to the service yet can create substantial burdens for patients; bringing this asymmetry into view is essential. So too is recognising that conventional productivity metrics actively count failure demand as output - until we distinguish activity that meets need from activity generated by a failure to meet need, headline measures will continue to overstate the value the system delivers.

However, the concept resists precise definition. What is not failure demand? Is all prevention failure demand? How do we distinguish failure demand from the inherent uncertainty of clinical practice? These boundary questions remain unresolved and may be irresolvable. We suggest this is not necessarily a problem: the concept's primary value may lie in its power to provoke investigation and conversation, rather than in serving as a measurement framework.

This report collates a fragmented literature and experiential knowledge base, offering a foundation for more informed discussion. The concept is not new, but it may be particularly timely given current pressures on NHS productivity and the limitations of conventional approaches to efficiency.

We offer a typology of signals of failure demand to provide a common language for recognising the diverse forms that failure demand can take. Our intention is to support practitioners, leaders and policymakers to make sense of what they are seeing, hold more informed conversations, and choose responses that fit the underlying problem. The typology acts as a starting point for more purposeful improvement efforts, rather than a rigid model to be applied prescriptively.

The value of the typology is in signposting appropriate responses for different signals. For relatively straightforward, within-service failure demand, the appropriate task is to diagnose and fix identifiable defects. For more complex, cross-boundary failure demand, the task is to create conditions for learning and adaptation. Treating an emergent, relational problem as though it were a simple technical one is likely to fall short. Conversely, treating a clear process defect as a complex adaptive challenge would be unnecessarily slow when the underlying cause is already knowable.

The role for policymakers and system leaders is to create the conditions for learning and experimentation. This means investing in systems thinking capability, resisting the temptation to turn failure demand into a target, and designing governance and funding arrangements that enable frontline teams to experiment with redesigning services around patient value.

Recommendations

For policymakers

Anchor the concept in value, not failure. Application of the concept should start from the question: “What is valuable to the patient/service user?” Failure demand is then understood as the gap between this and what actually happens – a framing that connects to existing NHS language around personalised care and avoids the defensiveness associated with the word “failure”. However, policymakers should resist extending the concept to encompass all preventable ill-health; a narrower, operationally focused definition is more likely to support meaningful action.

Use failure demand as a diagnostic lens, not a performance metric. Failure demand should be used to catalyse learning and local redesign, not to manage performance. Policymakers should resist creating targets or benchmarks around failure demand and instead give local systems the mandate and resources to investigate and experiment.

For system leaders

Invest in systems thinking and improvement capability. Addressing failure demand across services requires skills in systems thinking, operational management, human factors, and quality improvement that are currently in short supply across the NHS. Building this capability should be a priority, including within improvement teams whose analytical experience is often confined to specific pathways rather than whole-system analysis.

Use the signals typology to trigger learning cycles. Stakeholders recommended exploring failure demand in services that are both under significant pressure and known for poor patient experience, such as phlebotomy, outpatients, end-of-life care, and women's health. Integrated Neighbourhood Teams, with their focus on local responsiveness, offer a natural setting in which to embed failure demand and value as a core part of learning and improvement from the outset.

For researchers

Develop shared approaches to conceptualising and studying failure demand. The knowledge base on failure demand is fragmented across quality improvement, operations management, and health services research. Researchers should develop approaches that are sensitive to context and ambiguity, strengthening the evidence base while avoiding measures that could be misused as targets.

Explore the burden of failure demand on patients and carers. Patients experience the whole system end-to-end and bear much of the cost of failure demand, yet current metrics capture none of it. Research with defined patient cohorts, for example people with long-term conditions or learning disabilities, would help quantify its scale and impact, trace cascading effects across services, and make this burden visible.

Examine whether modern service frameworks might reduce failure demand. Models designed around continuity, relationships, and responsiveness - including integrated neighbourhood teams, comprehensive geriatric assessment, and personalised care - warrant investigation. Research

should examine whether they prevent avoidable downstream activity and improve outcomes for patients poorly served by conventional pathways, and under what conditions.

References

- Anderson-Wallace, M. and Downham, N. (2024) *Improving quality in healthcare: questioning the work for effective change*, London, Sage.
- Arabadzhyan, A., Castelli, A., Gaughan, J., et al. (2025) *Productivity of the English National Health Service: 2022/23 update*. York: Centre for Health Economics.
- Beacon, R. and Seamark, A. (2025) *Hospital of the future: ending the patient gridlock*. London: Re:State.
- Beardsell, I. (2025) "What if"-Two words that are breaking emergency medicine. *BMJ*, 390, r2051.
- Benniche, S., Everest, G., Hardy, S. and Gardner, T. (2025) *Mind the gap: public perceptions of the NHS and social care*. [Online]. Available from: <https://www.health.org.uk/reports-and-analysis/briefings/mind-the-gap-public-perceptions-of-the-nhs-and-social-care> [Accessed: 13/2/26].
- Berwick, D.M. and Hackbarth, A.D. (2012) Eliminating waste in US health care. *JAMA*, 307, 1513-6.
- Braithwaite, J., Glasziou, P. and Westbrook, J. (2020) The three numbers you need to know about healthcare: the 60-30-10 Challenge. *BMC Med*, 18, 102.
- CFE Research and Cordis Bright. (2023) *Evaluation of the Changing Futures programme - baseline report*. London: Department for Levelling Up, Housing and Communities, HM Government.
- Charlton, R. (2025). The Public Services Burning Platform: Using Existing Transactional Data to Make a Case for Relational Public Service Approaches. *Futures in Public Management*.
- Cook, J., de Lacy, J., Wood, J. and Mohammed, M.A. (2025) *10-Year Productivity Forecast for the English NHS: An Expert Elicitation Study*. [Online]. Available from: <https://www.strategyunitwm.nhs.uk/publications/10-year-productivity-forecast-english-nhs-expert-elicitation-study> [Accessed: 13/2/26].
- Cream, J., Wellings, D., Wenzel, L., et al. (2025) *Lost in the system: the need for better admin*. [Online]. Available from: <https://www.kingsfund.org.uk/insight-and-analysis/long-reads/lost-in-system-need-for-better-admin>.
- Dakin, F.H., Rybczynska-Bunt, S., Rosen, R., et al. (2024) Access and triage in contemporary general practice: A novel theory of digital candidacy. *Soc Sci Med*, 349, 116885.
- Dent, A., Plunkett, J. and Thomas, E. (2026) *Civic Rewilding - Applying the lessons from Big Local*. London: Kinship Works, Local Trust.
- Downham, N., Bricknell, S., Khan, S., et al. (2025) *General Practice — Prioritising effectiveness over volume?* : Develop Consulting and Cressbrook.

Fox, C., Hesselgreaves, H., Cooke, S., et al. (2025) *Policy Priorities to Support Relational Public Services*. Manchester: Manchester Metropolitan University.

Fozzard, K., Mu, Y., Elliott, J. and Rocks, S. (2025) *Can the NHS meet its 2% productivity challenge? Here's what experts think*. [Online]. Available from: <https://www.health.org.uk/features-and-opinion/blogs/can-the-nhs-meet-its-2-productivity-challenge-here-s-what-experts-think> [Accessed: 13/2/26].

Health Foundation. (2023, updated 2025) *Inequalities in life expectancy and healthy life expectancy*. [Online]. Available from: <https://www.health.org.uk/evidence-hub/health-inequalities/inequalities-in-life-expectancy-and-healthy-life-expectancy> [Accessed: 13/2/26].

Healthwatch. (2023) *Your experiences of getting prescription medication*. [Online]. Available from: <https://www.healthwatch.co.uk/blog/2023-08-31/your-experiences-getting-prescription-medication> [Accessed: 17/2/26].

Hensher, M., Canny, B., Zimitat, C., et al. (2020) Health care, overconsumption and uneconomic growth: A conceptual framework. *Soc Sci Med*, 266, 113420.

Hofmann, B. (2025) Addressing the overuse-underuse paradox in healthcare. *Med Health Care Philos*.

Hutchings, R., Davies, M. and Curry, N. (2025) *Falling through the gaps? A closer look at NHS continuing health care*. London: Nuffield Trust.

James, C., Denholm, R. and Wood, R. (2024) The cost of keeping patients waiting: retrospective treatment-control study of additional healthcare utilisation for UK patients awaiting elective treatment. *BMC Health Serv Res*, 24, 556.

Jones, B. and Pereira, P. (2024) *How improvement can help NHS productivity*. [Online]. Available from: <https://www.health.org.uk/reports-and-analysis/analysis/how-improvement-can-help-nhs-productivity>.

Kajaria-Montag, H., Freeman, M. and Scholtes, S. (2023) *Continuity of Care Increases Physician Productivity in Primary Care (June 05, 2023)*. INSEAD Working Paper No. 2023/23/TOM. SSRN.

Korenstein, D., Chimonas, S., Barrow, B., et al. (2018) Development of a Conceptual Map of Negative Consequences for Patients of Overuse of Medical Tests and Treatments. *JAMA Intern Med*, 178, 1401-1407.

Leary, A. (2024) "Rookie factor": why healthcare workforce policy needs to embrace experience. *BMJ*, 387, q2836.

Lowe, T. and Hesselgreaves, H. (2024) *Better outcomes for less money: the effects of working in a Human Learning Systems way*. Centre for Public Impact; Policy Evaluation and Research Unit; Civinet.

McDermott, O., Antony, J., Bhat, S., et al. (2022) Lean Six Sigma in Healthcare: A Systematic Literature Review on Challenges, Organisational Readiness and Critical Success Factors. *Processes*, 10.

Montori, V. (2020) *Why we revolt: a patient revolution for careful and kind care 2nd edition*, New York, Rosetta Books.

Moody, N. and Powell, T. (2025) *NHS Productivity*. London: House of Commons Library.

National Audit Office. (2025) *Lessons learned: Smarter delivery – improving operational capability to provide better public services*. London: National Audit Office.

OECD. (2017) *Tackling Wasteful Spending on Health*. Paris: OECD.

Opie-Martin, S., Cavallaro, F., Marszalek, K., et al. (2025) *One year on: is the government on track to meet its waiting times pledge?* [Online]. Available from: <https://www.health.org.uk/reports-and-analysis/analysis/one-year-on-is-the-government-on-track-to-meet-its-waiting-times> [Accessed: 13/2/26].

Palmer, S. and Torgerson, D.J. (1999) Definitions of efficiency. *BMJ*, 318, 1136.

Prendergast, T., Knight, H., Barr, B., et al. (2025) *The elective care waiting list: insights from linked data*. London: Health Foundation.

Radnor, Z. and Williams, S.J. (2025) *Lean as a Healthcare Improvement Approach*. Cambridge: Cambridge University Press. Available: <https://www.cambridge.org/core/product/A656E368A6DEFA5D3A2BBADB18B30C4>.

Randle, A. and Kippin, H. (2023) *Managing demand: building future public services*. London: RSA.

Reeves, C., Islam, A. and Gentry, T. (2024) *The State of Health and Care of Older People in England 2024*. London: Age UK.

Seddon, J., Hussain, I., Rubbra, T. and Wrighton, B. (2019) *Beyond command and control*, Oxford, Mayfield Press.

Shah, R. and Foell, J. (2024) *Fighting for the soul of general practice*, Bristol, Intellect Books.

Shanahan, P., Ollis, L., Balla, K., et al. (2021) Experiences of transition from children's to adult's healthcare services for young people with a neurodevelopmental condition. *Health Soc Care Community*, 29, 1429-1438.

Sinnott, C., Ansari, A., Price, E., et al. (2026) Experiences of access to general practice in England: qualitative study and implications for the NHS 10 year plan. *BMJ*, 392, e087367.

Sinnott, C., Price, E., Ansari, A., et al. (2025) What's been tried: a curated catalogue of efforts to improve access to general practice. *BJGP Open*, 9.

Taylor, B., Lobont, C., Dayan, M., et al. (2025) *Public satisfaction with the NHS and social care in 2024. Results from the British Social Attitudes survey*. London: King's Fund and Nuffield Trust.

Thomas, S. (2025) *Delayed diagnosis of cancer: a thematic review of general practice indemnity claims*. London: NHS Resolution.

Verkerk, E.W., Tanke, M.A.C., Kool, R.B., et al. (2018) Limit, lean or listen? A typology of low-value care that gives direction in de-implementation. *Int J Qual Health Care*, 30, 736-739.

Walley, P., Found, P. and Williams, S. (2019) Failure demand: a concept evaluation in UK primary care. *Int J Health Care Qual Assur*, 32, 21-33.

West, M. and Wallbank, S. (2025) *What does the NHS Staff Survey 2024 really tell us?* [Online]. Available from: <https://www.kingsfund.org.uk/insight-and-analysis/blogs/nhs-staff-survey-2024-tell-us> [Accessed: 13/2/26].

Wilkinson, E. (2025) NHS inflexibility is behind falling vaccine uptake. *BMJ*, 389, r1237.

The Strategy Unit

Email: strategy.unit@nhs.net

Web: www.strategyunitwm.nhs.uk

Twitter: [@strategy_unit](https://twitter.com/strategy_unit)

