locality

Saving money by doing the right thing

Why 'local by default' must replace 'diseconomies of scale'





VANGUARD

Locality, in partnership with Professor John Seddon of Vanguard Consulting carried out this groundbreaking research challenging the assumption that 'economies of scale' should be sought in the running of public services.

For their work on this report, we are particularly grateful to:
Neil Berry, Louise Winterburn and staff at Locality,
Locality's members who have contributed material to this study,
the Business Journalist Simon Caulkin,
Brendan O'Donovan and Charlotte Pell at Vanguard,
and all of the Vanguard consultants and clients involved
in the development of this material.

Locality is the leading nationwide network of community enterprises, development trusts, settlements and social action centres.
locality.org.uk

Led by Professor John Seddon, Vanguard Consulting helps service organisations change from a 'command-and-control' design to a 'systems' design.
vanguard-method.com

Published March 2014

Contents

Foreword

At a time of austerity cuts, mounting demand and rising expectations, the challenges to public services have never been greater. How we respond to these challenges will affect the lives of millions of people, and play a significant role in resolving the country's financial problems.

In recent years the prevailing view, in the Treasury and elsewhere, has been that public service efficiency can be driven by a combination of scale and standardisation.

That this strategy has experienced setbacks is not in dispute. Names such as Serco and A4E spring quickly to mind. But the government response has been to try to make the strategy work better, by enhancing the skills of government procurement teams, or improving supply chain management by prime contractors.

Meanwhile the underlying assumption, that the difficulties facing public services will be met through scale and standardisation, is not being challenged.

This report presents a counter view. We argue that scale and standardisation are the problem, not the solution.

As the report sets out, far too many public service systems 'assess rather than understand; transact rather than build relationships; refer on rather than take responsibility; prescribe packages of activity rather than take the time to understand what improves a life'.

The result is that the problems people face are not resolved, that public services generate ever more 'failure demand', that resources are diverted to unproductive ends, and that costs are driven ever upwards.

We all know there are countless examples of dedicated and skilled public servants (regardless of whether they are employed in the state, voluntary or private sectors). But the best public servants find themselves working against the grain, going against the system in order to do the right thing.

It is not public service which is at fault here. It is a system dominated by scale and standardisation. That is what needs to change.

In this report we have drawn on the private and public sector expertise and insight of Professor John Seddon and his team at Vanguard, as well as the on-the-ground experience of Locality's members and partners in communities across the country.

Our report sets out an alternative strategy. We propose that public services should be 'local by default', that they should help people help themselves, that they should focus on underlying purpose rather than outcome, that they should manage value not cost.

This, we believe, provides the best way to reduce demand, not amplify it, and to prevent problems arising in the first place, rather than accumulating costs which could and should be avoided.

Steve Wyler
Chief Executive
Locality

Executive summary

This report shows that the UK public sector is wasting millions of pounds on services that do not meet people's needs. When people's problems go unresolved, their needs remain the same or get worse, creating unnecessary demand and spiralling costs. The human cost is incalculable but can be felt by reading the true stories of Child A, Melvyn, Ruth and Jake in Part Ic.

The financial cost to the public sector can be measured empirically as the groundbreaking studies in this report show. By tracking multiple demands from individuals over time and across public services, it is possible to quantify the actual costs of a service from start to finish for each individual. Analysing hundreds of thousands of demands from many individuals makes it possible to confidently draw conclusions on where and how to reduce costs. If the experience in the few areas we have studied is typical, initial calculations suggest that potential cost savings for local authorities alone from a move to locality working could run to as much as £16 billion annually across England, with even further savings in other parts of the public sector.

This differs from previous studies of public sector resources because it starts from the service user and then counts every demand they make across organisational boundaries. The counting only stops when the original need has been met, crucially, as perceived by the individual, not by the organisation. It is also the first study of its kind to discriminate between artificial demand for public services, generated only as a result of an organisation not taking the right action, and the real demand experienced by the person who needs help. This artificial demand is called 'failure demand' ('demand caused by a failure to do something or do something right for the customer', John Seddon, 2003).

This report shows why public sector organisations fail to meet people's needs and why demand is rising. The two main causes, discovered empirically in the studies, are the belief in 'economies of scale' and the belief in the standardisation of services. Together, these beliefs prevent organisations from understanding and meeting people's needs.

Perhaps the most surprising finding, described in Part I, is that real demand for most public services is not rising. It is the artificial demand, created and amplified by organisations themselves which is rising. This finding marks a seminal moment in our understanding of demand for public sector services because it shows us exactly what to do. No further cuts or attempts to 'manage' demand by putting it online are required. We know how to reduce millions of pounds worth of unnecessary demand on public services; simply design services which are able to do the right thing for people in the first place. More effective services are more efficient, as people have their needs met more quickly rather than having to place numerous demands on the old unresponsive systems.

The effects of scale principles on the most disadvantaged and vulnerable people helped by local third sector organisations are described in Part Ib. The belief that 'economies of scale' are achieved by commissioning large public sector contracts has a number of damaging consequences with no increase in efficiency. One consequence is an increase in costly administrative burdens of tendering, compliance and monitoring, particularly troubling for third sector organisations, who strive to maximise resource allocation to the frontline and away from management and administration. More worrying is the impact on vulnerable people; they are provided with what has been commissioned rather than what they need. Other unintended consequences of large scale contracts include:

- the creation of silos and disjointed services across all sectors
- a decrease in competition and diversity of supply
- a decrease in innovation and cooperation
- an increase in uncertainty
- a culture of fear
- the erosion of independence

Taken together, this evidence represents a staggering opportunity for the UK to reconfigure public sector resources, saving the economy many millions, if not billions of pounds. The Vanguard Method achieves this empirically, starting with one person at a time, understanding their needs in context and building up a true picture of demand locally. As illustrated in Part III, this enables all public sector organisations in a geography to work together to design a bespoke, multi-disciplinary, evidence-based system that meets local demand. This approach, unlike many other attempts to join up services, does not require additional funding or encouragement from Whitehall. It does however, depend on the willingness of public sector managers to abandon unhelpful beliefs about 'economies of scale' and standardisation.

The principles and practice of this counter strategy are outlined in Part II, together with two case studies, one from the UK and one from the Netherlands. The example from the UK highlights the importance of understanding people and families in their own contexts and in their own language, away from standardised forms, scripted telephone conversations and official interview rooms. This approach shows the profound impact of helping people previously labelled 'troubled' and 'lost' to find ways of solving problems themselves. Not only does this approach improve lives and communities, it dramatically reduces future demands placed on the system.

The example from the Netherlands shows that understanding demand in human terms and providing the means for self-help are universal principles for effective and low cost services.

Part III describes the implications for policy and regulation. The report does not advocate further privatisation, nor conclude that private is good and public is bad. The conclusion is simple; if the public sector is to provide services that meet people's needs at reduced cost, scale principles must be abandoned.

The new principles for services that meet people's needs:

- are 'local by default'
- help people to help themselves
- ensure a focus on purpose, not outcomes
- manage value not cost

The report ends with a call to action. We know how to improve the lives of individuals and communities and the good news is that it doesn't take any more resources to do it. But it does take courageous public sector leaders who are willing to follow evidence and abandon old beliefs. Only they can do it.

Introduction

Public services work poorly. Every day the press carries stories of failure, waste and basic human needs unmet. Health, social care and children's services are buckling, while local authorities cut discretionary spending on libraries, parks and other local amenities to the bone.

At the same time the third sector, often referenced as part of the solution, is under ever increasing pressure as demand increases and funding declines, and may become part of the same crisis, its distinctiveness in danger of being progressively erased as it is co-opted into unreformed service delivery models by the public-sector regulatory and policy regime.

The third sector

The third sector – an umbrella term for social enterprises, voluntary organisations, cooperatives, charities, NGOs, civil society and community organisations – is important for public service delivery for a number of reasons. Distinct from both public and private sectors, its focus has historically been on the most disadvantaged and vulnerable, and third-sector organisations (TSOs) traditionally place high emphasis on social value. Less constrained by dogma and more inclined and able to work with their service users in the round, in these areas TSOs have shown considerable capacity for innovation. Although the sector is currently going through a tough time financially, overall its role in public service delivery is growing. This is a stated policy aim of government, but it is also happening naturally as TSOs are increasingly called on to deal with the knock-on effects of economic decline, cuts and the failure of public services to help those in need. Whether delivered by public or third-sector agencies, all services are now linked in a meta-narrative of despair: a public-sector doom-loop[1] in which rising demand meets finite or shrinking resources, leading to cuts and rationing into eternity.

But it is a crisis of our own making[2]. The news is full of stories of large-scale services of poor quality, which fail to meet demand and waste money. In some parts of central and local government there seems to be an unswerving belief that up-scaling contracts, vastly reducing the number of providers and defining a limited set of outcomes, will automatically lead to a reduction in the cost of services without a corresponding reduction in the quality of service. The commissioning and delivery of the Work Programme has clearly demonstrated the unintended consequences of designing a welfare to work programme shaped by a belief that the only way to save money and deliver services is by 'bulk buying' support, at scale and at rock bottom prices, with price being the main determinant in deciding who to award contracts to, instead of looking at what was actually needed and what works.

1. See this much quoted 'graph of doom': http://www.gmcvo.org.uk/graph-doom-and-changing-role-local-government
2. For an account of how public services have come to be in this state, see Seddon J 2008 'Systems Thinking in the Public Sector: the failure of the reform regime and a manifesto for a better way' Triarchy: Axminster

Unfortunately, the prevailing prescription for such failings is simply to increase the dosage (more with less, privatisation, 'better commissioning') and, as such, it offers a bleak outlook – as a wise person once remarked, doing the same thing over and over again with the expectation of a different result is a definition of insanity. Yet counterintuitively, the fact that current methods (mass-produced, silo-based, transactionalised services based on assumptions of 'economies of scale') are self-evidently wrong is a cause for optimism, not despair. As with any cul-de-sac, the way out leads in the opposite direction – in this case, creating services that treat people as whole individuals not parts, humans not transactions, and give them what they need to solve their problem, not the standardised package the system has specified in advance. As this report will show, far from being a luxury we can't afford, this is the only way we can afford it. As the examples demonstrate, when services are delivered this way, the vice of the doom-loop is broken. As problems are solved, demand stabilises, morale among both service users and providers rises, and costs drop out of the system.

Vanguard and Locality have been working together to examine this problem. Vanguard brings its experience of working with organisations in all sectors to transform service delivery, reducing total costs by solving people's problems. Locality and its 480 members work in some of the most deprived areas of the country, delivering a wide range of innovative and effective local community owned services. Together, we reject the misleading and flawed mantra that big services and scale are cheaper and more effective. Instead, we propose four principles that should guide future procurement decision making. Not coincidentally, these principles are the exact opposite of the scale dogma underpinning today's provision. Doing the right thing consists of providing services that:

- Are 'local by default'. Traditional scale economies are irrelevant and dangerous in services. Contrary to the present wisdom, what matters is not size but knowledge of context, and that can only be obtained on the ground.

- Help people to help themselves. Current services focus on needs rather than strength, as a result fostering dependency and increased demand. Human-shaped services build on strengths and promote responsibility instead.

- Are focused on purpose, not outcomes. Better outcomes are a consequence of effective intervention and thus cannot be managed directly. Management by purpose enables learning and improvement, as opposed to outcome-based management that drives dysfunctional behaviour, fosters cheating and hides failure.

- Manage value, not cost. Managing cost is at the heart of 'economies of scale', driving specialisation, functionalisation and an obsession with unit cost. Cost and outcomes-based management and payment by results are the main causes of the present system's dysfunctionality. It is methods and management focused on value and purpose that produces outcomes, not the other way round.

The net result of following these principles is to reduce demand instead of amplifying it. This is the key measure of achievement and the critical intervention point. Methods described here do not manage demand by rationing or passing it on to someone else. They reduce it – permanently – by solving the problem that caused it, preferably at first pass.

Designing services to meet people's needs and solve their problems rather than deliver standardised, mass-produced solutions reverses today's dynamic, breaking the tyranny of centralised scale thinking which has dominated public service delivery and holds it in the current impasse. Instead, following the principles that create economies of flow will ensure the continued social and economic legitimacy of local and third-sector organisations, currently in danger of being diluted under the prevailing assumption that 'bigger is better'. It is the missing link between 'local' and 'efficient' which explains how multiple small-scale interventions can be cheaper and better value for money than scale provision.

Economies of flow

'Economies of flow', in contrast to 'economies of scale', are what we create when we take a different approach to designing a service by following these principles:

- Any waste in a system represents, by definition, a failure to provide value for customers
- We should design to provide value, and that alone
- A focus on managing value will drive costs out of a system
- Control needs to be located where the work is done (controlled by the workers)

These principles enable the system to absorb the variety which it is presented with. See Seddon J 2003 'Freedom from Command and Control: A Better Way To Make The Work Work'. Vanguard Education Ltd., Buckingham.

Evidence base & methodology

This report features empirical evidence from two sources:

1. The most comprehensive study ever carried out on the demand placed on public services in the UK

The findings in this report are based on the cumulative results of hundreds of in-depth studies into hundreds of thousands of demands placed on the public and third sectors in the UK over the past three years.

Clients of Vanguard carried out the studies of demand across the following services, using the Vanguard Method:

- Local authority health and social care systems
- NHS hospitals, including demand into A&E
- GP surgeries and Clinical Commissioning Groups (CCGs)
- Third sector organisations
- Police services
- Fire and rescue services
- Care homes
- Housing services

The Vanguard Method

These studies followed a method developed by Vanguard – the Vanguard Method. It is an approach to improvement originally developed for private-sector service organisations and has the distinguishing feature that change starts by studying, obtaining knowledge of the 'what and why' of performance as a system (as contrasted with starting change with a 'plan'). In outline, the Vanguard Method provides the means to study services end-to-end, understanding service-user demands, following the demands through the services in order to understand how and how well the services work and identifying the system conditions that help or hinder achievement of purpose from the service users' point of view.

In every study, individual demands from service users were tracked through the system over time to understand how well the system understood and met their needs. For example, in one study, a total of 60,000 demands into a local authority adult social care service were tracked over the period of a year. In another study, hundreds of thousands of demands into a hospital were analysed over a two-year period.

The primary demand data was collected from case notes, database records, files, phone calls and other interactions with service users across all the organisations involved in each individual's case, for example, from local authority departments, the police, fire and rescue and other frontline delivery organisations in the public or third sectors.

All studies identified the barriers (or 'system conditions') that prevented the system from meeting the needs of the service user. The belief in 'economies of scale' was a predictable system condition in every study.

These studies have included the cases of many individuals, four of which (Child A, Ruth, Melvyn and Jake) have been chosen to illustrate particular aspects of the dysfunction caused by scale designs.

2. Research into 235 third sector organisations

Locality carried out a survey of 96 TSO leaders in the summer of 2013. Each organisation was invited to describe the extent and effects of scale ideas on their organisation. The survey was constructed and introduced in such a way as to ensure that responses could not be biased by an appreciation of the research interests or by the wording of particular questions. The results of this survey are detailed on page 24.

A further 139 organisational completions of Locality's 'Contract Readiness Checker' (see www.contractreadinesschecker.org.uk) were reviewed to understand relational issues of organisation and contract size.

Finally, the experiences of 9 local third sector organisations were followed up through in-depth case reviews. These case reviews were instrumental in forming some of the conclusions in this paper, and a selection of their stories are used throughout to illustrate certain key points.

Part I

The problem: 'diseconomies of scale'

a) How well do public services work for people who need help?

Studying the way services work for people in need reveals a disturbing story. In short, not only do they not help most people in need, they often actually make their lives worse. This is not because the people delivering these services, whether employed by public, private or third sectors, are not capable, but because of service design: a design based on industrial or 'scale' principles, at the heart of which is a focus on the management of cost.

What are the common features of these types of public services as they are presently delivered? What do they exist to do and how well do they meet that need? The answer to this question can only be found by studying what happens at the point where the 'demand' hits the system, and that is where we start.

Demand comes in many forms and walks in through multiple front doors

Analysing demand for these types of services, two things become immediately clear. The first is that the term 'service users' really means 'people who need help because their life has come off the rails in some way'. They include those who cannot attend to their own basic needs, whether because of mental health, drug or alcohol problems, and/or are experiencing relationship breakdown. The second is that people in need face a bewildering choice of venues to take their problems.

In one small town, an area of a few square miles, analysis revealed that demand could come in through any one of 130 'front doors'.

This reflects the fact that public services are highly functionalised and specialised. Front doors are maintained by local authorities, social landlords, police, government agencies, GP surgeries, hospital A&E departments, and third-sector services which are called on to fill in gaps in the service ecology. Some 'doors' are physical 'go to' places, others telephone service centres.

The needy can also present indirectly through the back door – by committing crime, self-harming or attempting suicide. Sometimes such cries for help are doubly indirect, as in those who commit a crime in order to go to prison to get help to kick a drug habit, for example.

The variety of demand

Listening to people's stories reveals a wide variety of demands: *'My relative is having trouble with basic care', 'I can't cope with my children', 'Please just take her into care' (mother of a 14-year-old), 'My son needs a statement' (of special educational needs), 'I want someone to care for me' (10-year-old), 'My wife has died and I'm lonely', 'I've been ill and need help to get back on my feet', 'I'm depressed because I've lost my job', 'I'm having trouble with my landlord'.*

Whatever the name on the front or back door, the most frequent demands on public services are related to ageing, drug or alcohol dependency, mental health problems and domestic violence.

Service 'consumption' (not necessarily the same as having a problem solved) displays a similar common pattern. In health and social care, 80 per cent of high consumers present for age-related reasons and 20 per cent with chaotic lifestyles associated with substance abuse. A very small proportion (<1 per cent) of high consumers are traffic accident or other high trauma cases requiring expensive ongoing care. High consumers represent about 1 per cent of the population but account for half of global consumption and 10-15 per cent of the total demand.

Common patterns underlie demand into other agencies. Thus in a housing organisation's 'rent arrears' function, 'rent arrears' was a catch-all for seven distinct personal issues:

- *I need help to manage my finances*
- *I need help to resolve my benefits issues*
- *I need help to get back to work*
- *I need help to move to a more suitable (affordable) property*
- *My relationship has broken down, I need help to cope on my own*
- *I need help to deal with my alcohol/drug problem*
- *I need to find affordable childcare*

Another housing organisation discovered that, of its demand into the 'allocations and lettings' function:

- *50 per cent of those presenting would never qualify for social housing*
- *35 per cent had problems that would not be solved by allocating housing*
- *Only 15 per cent were what might be thought of as in genuine need of social housing*

All of these studies of demand show that demands are person-shaped, not service shaped. They illustrate the importance of looking beyond the presenting demand to understand the context and underlying causes, i.e. to understand it in human terms.

At present there is no attempt to do this, since each service views citizens' needs through its own specialist lens. While, for example, people entering a benefits service front door will have a range of underlying problems around finance and debt, employment, housing and relationships, the service – encouraged to do so by policy and regulation – reduces the presenting demand to 'I want to claim benefits' and responds with 'If you can prove you are eligible I will process you as a claimant'. In the same way, many demands into health services are treated as medical, even though underlying problems are social in their nature. A specialised and transactional view of a citizen prevents service provision from being based on the contextual information that makes each demand unique.

This disconnect between service provision and need leads to two forms of sub-optimisation:

1. People's real, contextual, problems are ignored – people don't get what they need

2. Much of the work that is done is, ultimately, of no value, draining capacity from the system

The failure of service provision to match the variety of demand is an important theme to which we will return.

From demand to flow: studying citizens' demands through the system

Having identified the nature of demand, the next step is to study the 'flow' of work – what predictably happens to demand as it goes through the system. At a high level the flow can be described as 'assess–do–refer', in which 'assess' is any type of assessment to determine people's needs and/or eligibility for a service; 'do' is 'doing something', including dealing with the presenting demand, providing whatever the agency exists to provide, regardless of fit, or actually providing what is needed and solving the real problem; and 'refer' is referring the problem on to another agency.

Following the flow, it becomes apparent that people in need are frequently subjected to repeated 'assess–do–refer' cycles both within and across the various services (health, social care, benefits, housing, police, etc) that deal with them. It is also clear that most of these 'points of transaction' are in practice referral rather than action points – 'assess–refer' rather than 'assess–do'. An applicant rarely obtains a service at the first transaction point, instead being referred on to 'bounce around' from one agency to another until a decision is made.

One community health trust discovered that less than 1 per cent of demand was resolved at the first point of contact.

At most points of transaction staff record the contact, give advice or information or point the person to another front door.

An 80-year-old contacted his local authority asking for desperately needed respite care. He was sent a leaflet and told to get in touch with Age Concern.

A large proportion of demand comes in through service (call) centres where agents work to standardised scripts and processes focusing on whether the presenting demand fits with the services their organisation provides ('is this for us?'). If not the demand is screened out or referred on.

Screening out: criteria and thresholds

Each service is concerned with whether users meet its criteria for provision. Because managers believe demand for care services is rising, they focus on 'managing demand' – a euphemism for rationing or 'keeping people out'. 'Not for us' is one reason, the other is 'too low a level of need'.

In an adult social care service, 78 per cent of people initially referred for assessment were screened out. In a children's social care service, 82 per cent of the cases referred for initial assessment were closed at this point.

Assessment is seen as the key mechanism for rationing access to services, prioritising those most in need. In practice, its effect is to amplify demand rather than control it.

Repeated screening

If the demand is not screened out at first contact, it is forwarded on for assessment – in practice a second, more detailed 'screening' process, the focus once again being 'is this for us?' and if it is, 'does it meet our criteria?'

High volumes of re-presenting demand illustrate that demand does not go away because it is screened out of a system. When a person's problem isn't resolved (which is most of the time), he/she just re-enters the system through another of the bewildering array of doors a referrer can choose from. Decisions about where to refer applicants are dictated by what services exist rather than what individuals need. If a service has been commissioned that vaguely relates to presenting needs, people will be referred there. Each time a new service is commissioned it creates yet another referral door to add to the list.

The thinking behind referral is: 'this isn't for us, so it is up to someone else to help them'. What actually happens at the next transaction point is that the assessment (screening) process starts all over again.

It is striking that when people re-present to health and social care services they are treated and reassessed from scratch:

When leaders in a health system studied demand they were surprised to learn that 86 per cent of the demand hitting their system was from people already known to it.

Studying calls in an adult services contact team revealed that not a single demand was new: everyone contacting the system had been in contact before.

Because of this 'as-new', episodic view of demand, service users undergo an average of five 'assess–do–refer' cycles. Variation around this figure is enormous, with some people returning and being treated as 'new' demands a dozen times or more.

Demand is stable

The failure to provide services at the points of transaction has the effect of amplifying the volume of demand presenting to the system. Hence the refrain of constantly rising demand. Yet studies find that underlying demand into health and social care is both predictable and stable. What is rising is 'failure demand' ('demand caused by a failure to do something or do something right for the customer', John Seddon, 2003).

Re-presenting with the same problem is one type of failure demand. Other types include progress chasing ('the service hasn't been delivered', 'what is happening to my case?'), and re-work because of bureaucratic complication ('I don't understand how to fill in the form'). These create high volumes of administrative work, consuming capacity.

The extent of failure demand is enormous. Vanguard analysis suggests that it accounts for 80 per cent of demand into health and social care services, for example. It occurs not only within but across multiple services as people try to get their problems sorted.

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 the records of eight people 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.

In all, nearly 500 people were involved in the service-provider organisations, producing around 800 documents. The end result: in each of the eight cases, the presenting condition was either unchanged or had worsened and dependency had either remained stable or increased.

Failure demand does not just occur in primary services. It also creates pressures downstream in TSOs such as advice services:

Advice agencies: dealing with failure demand from other public services

Studying the work of member advice organisations in cities around the UK, Advice UK (2008[3] , 2009[4]) and Advice NI (2011[5]) found that 40-60 per cent of demand for advice comes from citizens trying to rectify mistakes or work around the failure of state or social housing agencies to provide effective service. The agencies generating most problems were the Department for Work and Pensions (DWP), Her Majesty's Revenue and Customs (HMRC), local authority benefits services and social housing providers. In one organisation advising social housing tenants, 95 per cent of failure demand was caused by DWP. Advice costs generated by these failures are conservatively estimated at £500m a year, while the costs to DWP and HMRC of re-work and legal appeals will be much higher: one study found that 90 per cent of cases are eventually won by the appellant. The report's authors suggest that the increasing trend to fund advice organisations solely for advice transactions obstructs opportunities for learning as well as discouraging broader cooperative work to tackle waste and improve services.

The effect of the rationing system is to make those in need keep presenting (creating demand that is not going to be satisfied) until their problem becomes serious enough that they meet the assessment criteria and can be 'screened in'. This can take many cycles and sometimes years.

In one case, a family that first sought help from its local authority in 1995 was finally screened into the system more than 13 years later.

Some people only present because they have reached a crisis point.

The invisible problem of failure demand

Failure demand is the result of the failure of provision to match the variety of demand, itself the product of scale thinking. It is a 'diseconomy of scale', a consequence of the way work is designed and measured. The principal causes of failure demand are:

Managing cost

The starting point for most service managers is cost. Cost is assumed to be a function of scale. Scale is delivered by specialisation and standardisation, which has enabled the massive growth of outsourcing.

3. Advice UK (2008) 'It's the System, Stupid! Radically Rethinking Advice' AdviceUK: London see http://www.baringfoundation.org.uk/ITSS.pdf accessed 7/8/13
4. Advice UK (2009) 'Interim Report: Radically Rethinking Advice Services in Nottingham' AdviceUK: London see http://vanguard-method.com/images/R/DX/455-NottinghamSystemsThinkingPilotInterimReport.pdf accessed 7/8/13
5. Advice NI (2011) 'The Big Idea: Putting People First' AdviceNI: Belfast, see http://www.adviceni.net/publications/PDF/Systems per cent20Thinking per cent20Report per cent20the per cent20big per cent20idea per cent20March per cent202011.pdf accessed 7/8/13

But as we have seen, specialisation and standardisation lead to services that match the convenience of the commissioner, or sometimes the provider, not the variety of the need. For some users the service is overspecified, for others underspecified and for still others irrelevant. When people don't get help that matches their need, they re-present or present to different services or TSOs until they do. Budget management, screening and other attempts to manage cost by rationing have a similar effect. For example, raising the threshold of need to qualify for care assistance means that many people who could have stayed in their own homes with small domestic adjustments (most commonly walk-in bathing facilities) end up in expensive care homes where they don't want to be. Managing cost generates failure demand which drives cost up.

Activity-based performance measures and targets

In specialised, functionalised organisations, ultimate purpose is invisible so performance is driven by intermediate targets ('pick up the phone within three rings', 'reply within three working days', 'complete 90 per cent of assessments within two weeks'). The de facto purpose then becomes 'do the assessment' and 'tick the box'. In all kinds of care services, the result is that firstly many people undergo assessments that they don't need and which do not lead to any service outcome; and secondly that in acute cases, some assessments are undertaken just to make a referral. At the first appointment with the referred organisation, the assessment is repeated.

In a striking example of the target becoming the purpose, in the large majority of cases the assessment activity itself becomes the service 'response' – the 'do' step. Assessments are completed only so that people seeking help can be referred on, so that arbitrary review deadlines can be met, or so that the case can be closed.

In one police force, 68 per cent of assessments of people labelled as vulnerable were simply logged and filed. Some 87 per cent of those assessed re-presented back into the police on average an astonishing 17 times.

In adult social care, many people are sent to care homes because of the pressure on social workers to close cases. Once in a residential home, a patient can be deemed stable enough to come off the worker's caseload.

Because targets and performance management focus on the unit costs of activities, not the end-to-end cost of service provision, they do not discriminate between failure demand and real demand, which all becomes 'demand to be managed'. Failure demand is the most important source of waste in public services and correspondingly is the most important lever for improvement. Targets make it invisible.

Risk management

Growing emphasis on risk management (itself an attempt to mitigate the unintended consequences of a dysfunctional system) also has a powerful effect on assessment and provision of services. Meeting assessment targets creates de facto purposes such as 'keeping the service providers safe and not at risk of going to court' or 'keeping our institution safe from bad publicity and our officers safe from possible disciplinary action'; providing services as specified is a means of demonstrating 'accountability' and limiting perceived risk.

A study of eight people with drug or alcohol dependency showed they presented to GPs a total of 124 times; the system carried out 4,300 activities, creating 800 documents. Just 10 per cent of the activities were related to helping them, the remaining 90 per cent relating to approvals, reporting against targets and accounting for performance to commissioners. None of the cases improved.

What 'safe' looks like is defined by the system, for the sake of the system. For example, a person may be medically fit to leave hospital but deemed not safe for discharge until the system has made them safe by prescribing equipment or a specific level of care – this is done without knowledge of their need in context and generates wasteful activity.

The deadening effect of risk management increasingly affects regulated TSOs, which are often prevented from delivering what they know is really needed, and over time can themselves become overly cautious, both to the detriment of service users' quality of life.

An established and respected TSO in the Midlands is embarking on a new affordable housing scheme for their neighbourhood. Although it is a new venture for the organisation, its systems and financial plans were deemed sufficiently robust for a bank to offer to finance the £4.2m scheme. However the scheme required the local authority to release the land for development, and it was concerned about risk. It therefore insisted on delays and forced the TSO to work in partnership with a housing association to bolster its management capabilities. The TSO is still ultimately driving and responsible for the development, but the project is happening a year later than originally planned, and the estimated cost has now grown to £6.7m, with the additional costs picked up by the public purse.

The consequences of risk aversion are delays in service provision, people staying in hospital longer than they need to, domiciliary care being interrupted whilst a new care provider arranges for a risk assessment to be carried out, providing services to meet specifications rather than what's needed.

A victim of repeated domestic abuse stopped asking for help because each time she called the police they simply filled in the same standard 27-question risk assessment. This triggered a referral to social services and a risk assessment for her children, but produced no result for her, the abuse and suffering continuing as before.

Outcomes-based commissioning

Commissioning represents scale economy arguments applied to procurement. The philosophy is unambiguously to create markets for public services by documenting needs, drawing up specifications and tendering for provision. Outcomes-based commissioning (often using PbR mechanisms) goes further and aims to secure best value through price competition and transfer of risk to the provider through contracts under which providers only get paid if they deliver the results specified. This type of commissioning frequently favours large organisations which have the financial muscle to shoulder the financial risks and cash-flow requirements associated with PbR contracts, particularly for major central government schemes such as the Work Programme, where single contracts cover whole regions.

The consequences for people who need help

Failure demand is the symptom of a system that is unable to understand people in context or respond to their real needs. In those circumstances, people quickly learn that when they ask for help what they will get is assessment and referral. As a result some stop asking for help (a 'success' in the current system). Others resign themselves to accepting what's on offer even though it doesn't help. Most comply with the reviews and assessments which the system generates but which they haven't asked for to get what they can. Some feel so overwhelmed by the many professionals now managing their life that they give up trying to help themselves. Such users are usually labelled difficult or non-compliant and sometimes visited with sanctions or refused

further service. The overall consequence, however, is the same – a failure to solve the original problem, which as a result becomes worse or more complex. People with problems continue to place demands on, or be referred between, multiple services, inevitably consuming more resources across the system as a whole.

In 2006 eight-year-old 'A' was referred to social care by his primary school because he was a victim of domestic violence.

In the same year A's mother requested he be statemented[6] because of behaviour concerns, but this was not deemed appropriate. Since then A's troubling behaviour has escalated in both frequency and seriousness. A decision to statement him was taken in 2011. He has moved school five times in six years.

Now 15, A is heavily involved with the criminal justice system. He currently has five convictions, including for vehicle crime, robbery and possession of drugs. He was arrested twice in 2012 and is now excluded from his special school.

Over many years and several generations A's immediate and extended family has been well known to a large number of statutory and voluntary agencies. Domestic violence has been a frequently recurring theme.

Systems that fail to help

Today's public services are not designed for 'people who need help'. In the manner of a hospital set up to deliver a specific intervention – a replacement hip or cataract removal – they are designed to batch-process fixes for predefined one-off issues and then close the books. In consequence they are systems that assess rather than understand; transact rather than build relationships; refer on rather than take responsibility; prescribe packages of activity rather than take the time to understand what improves a life. As in any system that fails to solve the underlying problems, they amplify work, appearing frenetically busy while accomplishing less and less. Based on identifying needs rather than strengths, they fail to help individuals and communities build self-sustaining support systems that increase agency and independence, instead increasing resource consumption and dependency and accelerating decline.

These are systems that obsess about cost, yet paradoxically drive costs up. Screening, assessment and gaining approval to provide services are all capacity-hungry activities in themselves.

In health and social care, attending to these processes consumes around 75 per cent of front-line practitioners' time, rising to more than 90 per cent for middle and senior managers.

The activity of protecting cost is itself costly, with rationing and pot-juggling not only absorbing capacity but creating delays. Delays and rationing lead to an escalation of people's problems, consuming still more resources. Finally, the more public-sector services managers seek to contain apparently inexorably rising demand through thresholds, budgets, aggregation and centralised commissioning, the greater the volume of failure demand – in other words, the greater the 'diseconomies of scale'.

6. A statement of special needs is a formal document detailing a child's learning difficulties and the help that will be given.

‘System conditions’ at the heart of the failure to perform

Specialisation:
the reason for many ‘front doors’, the focus on fitting need to provision rather than providing to need; resulting in multiple agencies involved in service provision.

Budget management:
the driver for ‘managing demand’ (keeping people out).

Standardisation and commissioning:
resulting in services failing to meet individual needs.

Activity measures/targets:
resulting in assessments completed but people being referred on, creating an episodic approach to people.

Lack of continuity in relationships:
the consequence of taking an episodic approach, losing knowledge of people’s needs and resulting in duplication of activity.

Thresholds and criteria:
resulting in people being turned away.

Together, these system conditions are what might, for brevity, be labelled as ‘scale’ thinking; all have their roots in the economic propositions that economy is achieved through scale and competition leads to efficiency.

b) Scale principles and civil society

As we have noted, the third sector has historically been focused on the most disadvantaged and vulnerable, placing high emphasis on social value and having a reputation, among the best, for innovation. Successive governments have embarked on a series of initiatives designed to extend scale thinking to the third sector.

At present, the more the third sector is brought into formal public service delivery roles, the more it is subject to the policy and regulatory conditions that have shaped the experiences of public services described in Part Ia of this report. Central to these policies is the achievement of scale economies through commissioning (see Part Ia).

'Commissioning is a ... set of service delivery processes which involve consultation, needs assessment and service planning and design... commissioning is about deciding what to buy and how[7].

For TSOs and the communities they serve the recent shift to large-scale commissioning in a number of key areas of service delivery has had particularly serious consequences. As well as eroding much of their cost effectiveness, in constraining TSO autonomy commissioning also limits their ability to provide alternatives. In a survey of 96 TSO leaders by Locality in 2013, respondents report that current approaches to commissioning and procurement:

Increase administrative burdens, particularly around tendering, compliance and reporting

A TSO in the North West was an experienced and successful business support provider but recently found itself at the bottom of a long supply chain where 'all the costs were sucked into audit and compliance'. Three separate organisations (the original commissioner plus two layers of private sector contract managers) subjected the TSO to three separate but similar audit processes. Overall unit costs for the service were much higher than previously, but the organisation only received 40-50% of those unit costs to actually deliver the service, with the remainder invisibly swallowed up through complex management chains above them. Its much reduced front-line service also became bureaucratic and impersonal as a result of new contract terms and inadequate resourcing. 'Everything is process now... It's a cold-war mentality – the main preoccupation of commissioners is that someone, somewhere, is screwing them.' The organisation could no longer cover its cost of delivery within this new management model and ultimately had to close its business support operation down.

Create silos and disjointed services

Many TSOs now have to work to multiple contracts, because commissioners divide and subdivide services by specialisms (e.g. parenting programmes, home support, family intervention programmes, etc.), thresholds, age ranges and geography, leading to a need for increased coordination between service providers and internally to provide joined-up service to users. The up scaling of contracts is leading to silos and disjointed local service provision, with previously co-ordinated and seamless service delivery within local communities being dismantled, opening up gaps, inefficiencies and failure.

7. From Macmillan R 2010 'The third sector delivering public services: an evidence review' Third Sector Research Centre Working Paper 20

A TSO in the South West was able to integrate youth provision into a seamless, family-oriented approach, that starts with the birth of a new baby, continues through childhood, teenage years and into adulthood – where a range of adult services including those for the elderly are provided at the heart of the community. Workers at the TSO were able to develop a real knowledge and understanding of each family and build on their strong relationships with the family and individual children, to prevent issues arising or quickly spot signs of children at risk. When the LA decided to move from working with 65 local providers to nine large contracts, the TSO lost the contract to provide youth and play service, creating a gap in provision at a key transition point in children's lives. Vital information and deep-rooted relationships with children and their families are now lost. Under the new provision fewer activities are taking place, and the closely related web of informal local support has unravelled. Young people are 'hanging around' with nothing to do, while a whole range of excellent local providers have been excluded from consideration by size and bidding conditions, existing community services have stopped, and for many their continued existence is now in doubt.

Increase risk

With management costs and profits of non-delivery organisations further up supply chains absorbing significant proportions of public service budgets, third sector suppliers are often required to deliver inappropriate and uneconomic frontline services. This threatens the survival of important suppliers and/or reduces their potential for future growth. Business models are high risk and low/no margin, and this is currently having a significant inhibiting effect on the development of the emerging 'social investment' market.

Fail to meet real needs

In practice, commissioning means people are provided with what has been commissioned rather than with what they need. Because commissioners have no reliable method to determine need in a geographical area, the size of commissioned 'lots' often fails to reflect an understanding of local demand, meaning that the public purse has to carry the cost of over-provision – giving people more than they need – under-provision – giving people insufficient to meet their need (creating failure demand) – and providing some people with services that create no value at all.

A community association with niche expertise in services for complex domestic violence cases failed to win a new, aggregated borough-wide tender. The association estimates that 25 per cent of its caseload was complex – the needs of clients could not be met by telephone support or a visit to association offices (the location of meetings is vital in cases where a client is closely monitored by their violent partner). The winning contractor did not include in-depth support for complex cases in its service provision – clients must use a telephone help-line or visit the provider's office. However, the very real need for this type of support still exists, so the organisation is doing everything it can to continue its work with some of the most vulnerable clients, even though it no longer receives funding from the local authority.

Increase uncertainty

With contracts sometimes running for shorter periods than conventional grant-funded arrangements, some TSOs are experiencing higher turnover of staff and fear a loss of intellectual and social capital.

'The TUPE process is a considerable burden... A major concern is the loss of intellectual capital. TUPE potentially means that staff transfer from one provider to another, taking their knowledge and experience with them. This could help their new employer win further contracts at the expense of the employer who trained and developed them.'

Diminish innovation and cooperation

Standardisation, adherence to contractual terms and 'best practice' inhibit innovation which, ironically, previously was the hallmark of the TSO sector. Competition also has the side effect of discouraging sharing and collaboration amongst TSOs.

'Before the tendering process, there was a level of trust between organisations. Of course there was ongoing competition, but dialogue was reasonably open and there was some collaboration and joint working. All this has gone. The climate of suspicion means that there is less joint working both strategically and on the ground. The other consequence is that there is now no strategic dialogue with the local authority.'

Promote a culture of fear

Some TSOs may withdraw from lobbying and advocacy[8] for fear of having their funding affected. In the relationship between TSOs and commissioners, the latter now have all the power.

The third-sector verdict

Responding to a Locality survey:

- 80 per cent of TSO respondents reported that larger contracts had reduced or were set to reduce their opportunities to provide services
- 81 per cent said that larger contracts had diminished or were set to diminish both the range and quality of local services
- 85 per cent believed that larger contracts would not increase efficiency
- 26 per cent of respondents estimated that administration and compliance costs on contracts accounted for more than 20 per cent of the contract value

Barriers to winning public-sector contracts

The focus of most efforts to increase the role of the third sector in public sector service delivery is on 'contract readiness'. The assumption is often that third sector organisations lack the breadth and depth of delivery skills and track record, and lack organisational systems and management skills. The Locality survey suggests that these assumptions are to a large extent misplaced:

8. Alcock P, Butt C and Macmillan R 2013 'Unity in Diversity: what is the future for the third sector?' Third Sector Research Centre

Main barriers to winning public-sector contracts were reported by TSOs to be (in order of importance):

- size of contract (53 per cent)
- bureaucracy/administration (44 per cent)
- excessive risk in contract terms (36 per cent)

Skills (14 per cent), organisational capacity (19 per cent) and price competitiveness (17 per cent) were far less regularly cited as barriers.

In short, the results of imposing conventional industrialised thinking on the third sector are the same 'diseconomies of scale' that drive the public sector doom-loop. The cost of 'best value' is services that reduce unit costs but drive up the real end-to-end cost of provision. The cost of transferring risk is fake results (cheating), increased bureaucracy and loss of TSOs willing and able to provide good local services for the communities. The cost of size is greater administration and coordination costs, loss of local knowledge and commitment and ability to learn. The cost of all of them combined is failure to meet needs which fuels failure demand, the unseen vampire sucking more than half of all resources and capacity out of the system. By jeopardising the comparative advantages of TSOs, cutting their capacity for innovation, limiting the scope for developing intelligent partnerships, reducing their scope in service provision, driving some out of business and demoralising TSO personnel, scale thinking in general, and commissioning in particular, has the effect of undermining the very qualities of intensely local connections and commitment which both attracted knowledgeable local volunteers and made TSOs appealing as complementary partners to public-sector agencies in the first place.

c) The individuals' stories

The following case studies illustrate what actually happens to people when their lives go off the rails and they seek help from public services. They have been chosen for being representative. The cases were compiled by studying records in the various agencies over time.

Child A

The earliest recorded contact with child A was in 1997, when the child's mother contacted the council's children services department four times, saying 'I need help with my young son', then four, who was displaying inappropriate sexual behaviour. All contacts were recorded as No Further Action (NFA).

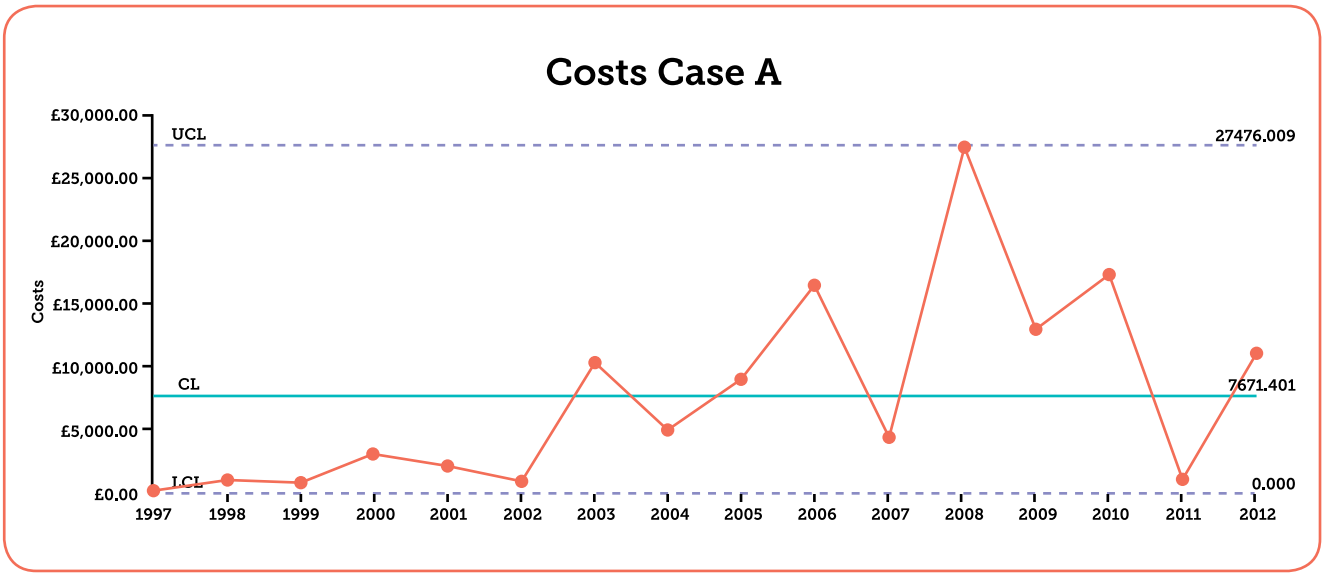
In 1998, A's mother asked her GP for help. The GP made a referral to child adolescent mental health services. In 1999, a child protection medical took place and A was briefly placed into foster care. In 2000, after a visit by social workers the boy was referred to an agency specialising in children's mental health. In 2001, the family was moved into new housing. A visit to the hospital resulted in a referral back to the children's services department. In 2002, an emergency out of hours contact was made following the child's display of abusive behaviour, with child A referred to another mental health agency. A formal Initial Assessment was started by children's services, four years after the mother had first contacted the department. In 2003 the mother began a course of treatment for alcohol problems. In 2004, a 999 call was made to the police following three separate domestic violence incidents.

A's mother had a second child. A childminder was assigned to the family for 15 hours a week and the family was referred to the local SureStart centre for support. In 2005, the police took another 999 call because of domestic violence/alcohol abuse. Child A was reported as missing on five separate occasions. At this stage, the children's services department began to take greater interest in the case as he was now deemed to be sufficiently in need to merit serious attention. He was referred to another agency, which specialised in working to build the self-esteem of young men. An alcohol worker was assigned to the mother. In 2006, there was a 999 call to the police plus one missing person report.

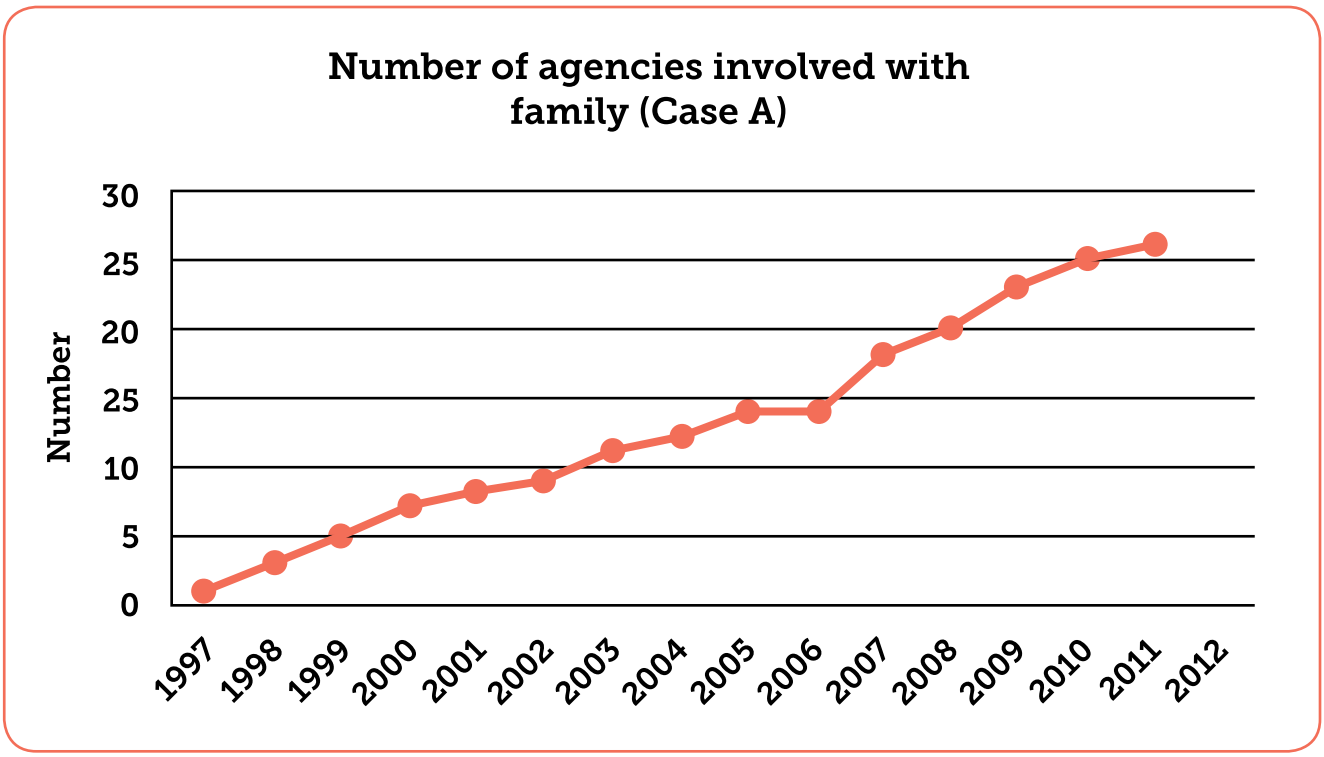
In 2007, after 11 unauthorised absences A was excluded from school. Another missing person report was filed. By 2008, the school was expressing concerns about the child. Three missing persons reports were recorded, with the child absent for a total of 12 days. He was arrested. At this stage, the family's second child was referred by children's services to begin an 'Early Years Intervention' because of behavioural issues. In 2009, the mother visited the children's services office to ask for help. Police reported disturbances at home and child A went missing another six times. He was sent on a residential trip by children's services, and had a laptop donated to him. Child A was placed in foster care when the school reported its concerns. At this stage, child A committed a serious crime (a sexual assault) and then broke his bail conditions.

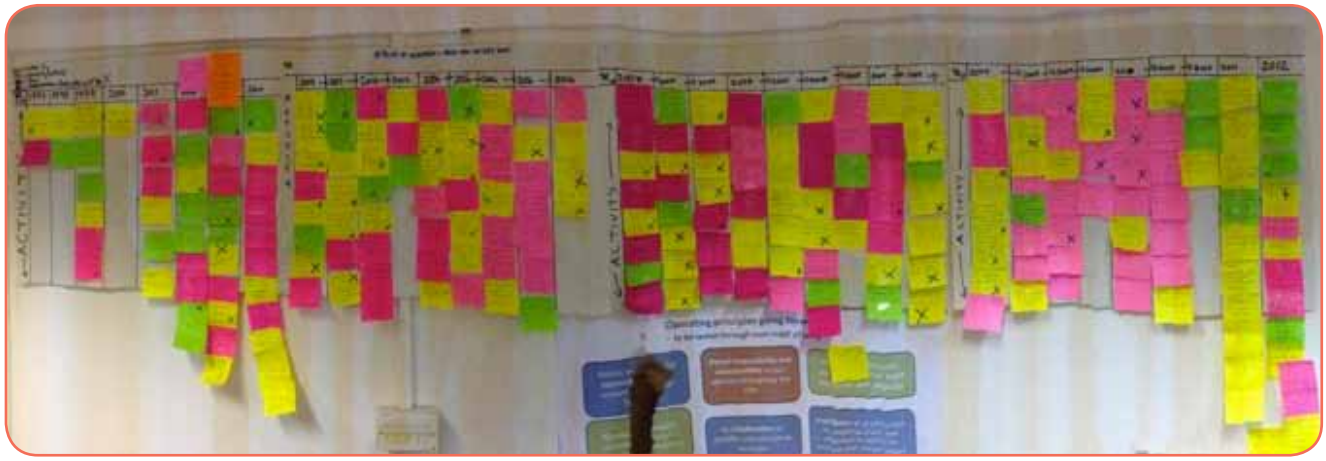
In 2010, A received a non-custodial sentence with Youth Offending Service (YOS) support. The service reported non-engagement with YOS activities and the boy was found to be in breach of his referral order. He returned to the magistrate's court where the mother was ordered to attend a 12-week parenting programme and the boy was referred to a psychologist. In 2011, concerns were expressed to children's services by the wider family about the youngest child's behaviour. The office sent social workers to visit the mother on three separate occasions. In 2012, by now moved into a hostel, A was reported missing on 18 occasions. By 2013 the authorities had lost track of his whereabouts, but he was presumed to be homeless.

The graph shows estimated yearly costs of interactions with the agencies involved. These totalled £130,000, the sole result being that the family's situation deteriorated over time.



The number of agencies involved with the child's wellbeing continued to grow, as the graph below shows.



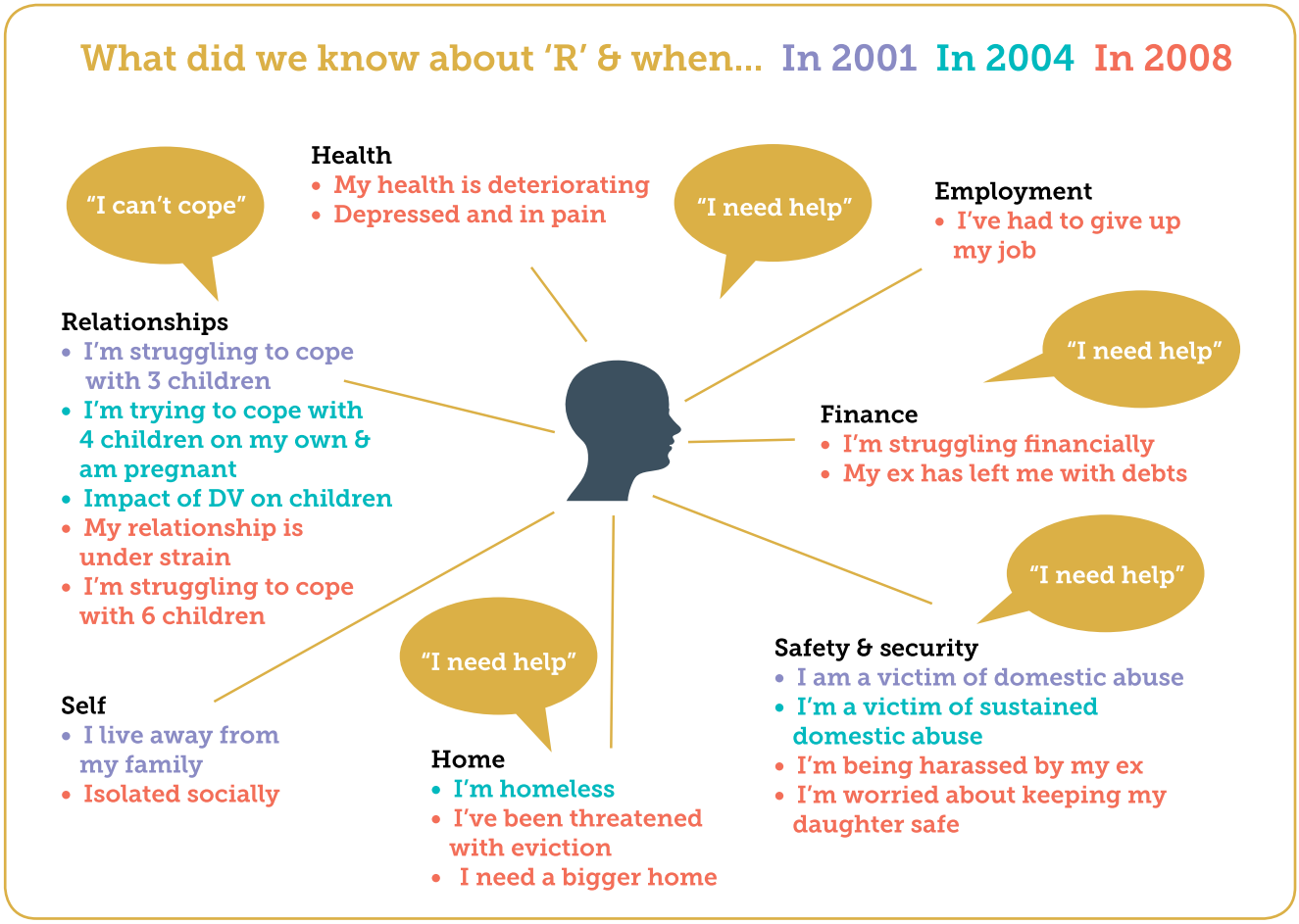


This picture shows all of the records made for this case by the various agencies on their IT systems. Each different colour represents a different agency's involvement.

The case features many of the characteristics described in Part Ia. When the mother first requested help, the system judged her needs insufficient to warrant action. Further requests for help attracted low-level standard packages of support. Only six home visits were made during the whole process, although absence from school was a recurrent problem. There were repeated referrals from one agency to another and no attempt to build a relationship with either mother or child in need. Many opportunities to change the direction of travel and avoid both greater dependency and increased costs to the public purse were passed up. The situation deteriorated until more costly interventions (in the criminal justice system) were required.

Ruth

Ruth was a victim of domestic violence. She has six children with three different fathers. While the first contact with Ruth was in 1996, the team studying cases built a picture of the transactions between Ruth and public services during 2001, 2004 and 2008:



Ruth's problems began in 1996 when her first husband was abusive. To escape the abuse she moved. The same happened with two subsequent partners; moving around caused problems with housing benefits and council tax. Calls to police resulted in form-filling, but no action. It was only when the children started running away and skipping school that the police took interest, but only to refer the case to social services which threatened to label the children as 'at risk'.

Ruth had by then stopped working to be with the children. Although she was also experiencing symptoms of a degenerative illness for which the NHS supplied medication, she was judged below the needs threshold for access to social services.

When her first partner reappeared demanding to see his daughter, Ruth feared further violence and was concerned for the child because her father was accused of a sexual offence. Calls to the police again led only to form-filling.

Social services commissioned a Family Intervention Programme resulting in a series of assessments. These led to two of the boys being sent twice to the same anger-management course and Ruth two identical Parenting Programmes. She was also sent for assessment for a Promoting Independence Programme, for which she had been previously assessed and refused; she was refused again for failing to meet the criteria.

As her health deteriorated Ruth became unable to climb the stairs. The children took advantage to run amok. Unable to take her children to school, Ruth requested a wheelchair which was refused because she didn't meet the threshold. Sanctions were placed on the children for truanting. Ruth borrowed money from friends and family to buy her own wheelchair to accompany the children to school. Social workers described Ruth as 'deviant' and 'working the system'.

When children's social care made a referral to adult social care, Ruth was referred for a bathing assessment (could she bathe the children properly?) which resulted in her being given a bathing stool. Needless to say, the bathroom was upstairs. Ruth was at risk of losing her tenancy because of the state of the property.

During a dispute with a neighbour the police were called and removed the children.

What Ruth said she wanted:

- 'I need help with housework and...'
- '...gaining access to the first floor of the property.'
- 'These two things would have such a profound effect on mine and the children's lives.'

What Ruth received:

- Two anger management courses for two boys
- Two parenting programmes
- Help cleaning one bedroom
- Toilet frame, perching stool and bath board for a bath she could not access
- Family intervention programme

And it took this many people to deliver it...

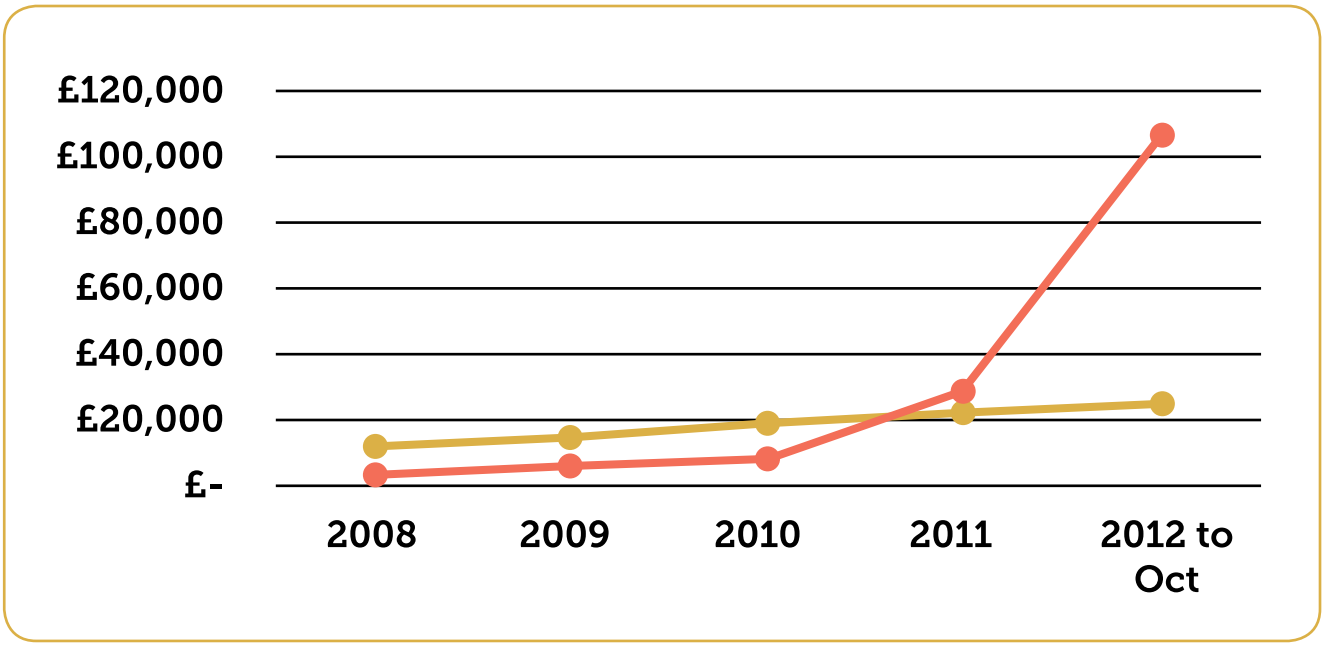
- Eight social workers
- 22 support workers allocated
- 30 referrals in core flow
- 16 assessments in core flow
- 36 teams/services

Cost of what Ruth wanted

Cleaner, assume 10hrs/wk for four years	£ 14,560
Move to suitable property (current home unsuitable for adaptation)	£ 1,200
And/or stair lift	£ 5,000
Total	£ 20,760

Cost of what Ruth has received since 2008, based on costs as at 2009: **£106,777**

The graph below shows cumulatively how the costs of the case have risen over time. The lines show estimated actual cost (red) compared with the cost of what Ruth needed – a difference of £86,000:



Between 1996 and 2012 Ruth experienced 129 different interactions with public-sector agencies. From running a successful business with her first husband she deteriorated to the point where she was plagued by ill health, had her children removed and was entirely dependent on the public purse.

Ruth has since been taken on by one of the new 'Wellbeing' teams piloting locality working (see Part II); she is now in suitable accommodation with her children and her situation has stabilised.

Melvyn

Melvyn, 75, is an ex-miner who lives alone in a council-owned bungalow. All his life he has suffered from epilepsy which interrupted his schooling with the result that he is unable to read or write. He has an ongoing lung condition (COPD[9]). Melvyn takes medication to control his seizures and uses a nebuliser to help with his breathing. He also has a history of urology problems. Melvyn brought up his three children after his wife died when the children were still quite young. His brother Graham and sister-in-law Mary provide support by accompanying him to appointments, dealing with all his correspondence and acting as the main contact for services. Graham is 85 and feels a great sense of responsibility for his younger brother.

What matters to Melvyn?

- I want to go out and about
- I want to be free from pain
- I want people I'm familiar with helping me
- I want to be in control of my life and make my own decisions
- I want to stay living in my home

How did the system respond to what mattered to Melvyn?

'I want to go out and about'

Melvyn enjoyed going out. He had a lady friend; they liked to walk her dog together and he looked after the dog when she was away. In July 2011 Melvyn spent 12 days in hospital with an aggravation of his COPD (a preventable admission). The first thing he did on returning home was to walk the dog. Unfortunately he had a fall and went straight back into hospital. Despite being medically fit within four days he did not return home until 47 days later. Delays were caused because the system judged him not 'safe for discharge' and he needed to wait for assessment and the subsequent care package to be set up. Melvyn caught a chest infection whilst he was waiting to go home. On discharge he was told it was not safe for him to walk unaided or go out by himself and that he should no longer have the dog to stay at his home. He was discharged with a perching stool, raised toilet seat, commode and a four-times-a-day care package for prompting the taking of medication and personal care. Before his hospital stay, Melvyn received low level support (meals on wheels) and occasionally used a walking frame.

Following instructions that Melvyn must not go out alone, his sister-in-law requested a ramp suitable for a wheelchair. This was agreed after assessment, but it took seven months to arrive. During this time Melvyn regularly told care workers and nurses that he was fed up and bored. He had several (mostly preventable) falls within his home triggering five falls assessments which led to no action - they simply fed the CQUIN[10]

9. Chronic obstructive pulmonary disease
10. CQUIN stands for Commissioning for Quality and Innovation. This is a form of Payment by Results (PbR) in the health service.

measures and were filed. Meanwhile Melvyn became very anxious about falling and one year on from regularly walking the dog would no longer venture out of his home.

'I want to be free from pain'

Melvyn had a permanent catheter for a number of years and operated a flip flow catheter himself without difficulty. In September 2011 there was a problem with leakage and a decision was made to switch to a catheter with a leg bag. This was the start of an unresolved catalogue of problems as a result of which he spent most of 2012 in pain from infections and blockages. This resulted in one hospital admission for 42 days and repeated visits to A&E for re-catheterization. District nurses visited at least weekly and re-catheterized as necessary. The community matron made monitoring visits. Leakage was still a problem and infections continued. When he was in pain, Melvyn used his Telecare service, often several times a day, to ask for help. In addition, it was now more difficult for Melvyn to manage the walking frame with the catheter bag and he fell several times. The community physiotherapist referred him for 'gait training'.

Melvyn spent more and more time sitting in his chair. It was electric and operated via a remote control. Melvyn often needed to summon help from an ambulance crew or Telecare responder, particularly at night, because he had become 'stuck' in his chair. The real problem was that Melvyn could not read the buttons on the remote control and gets confused how to operate it. No-one understood this and no-one had built a relationship with Melvyn whereby he would feel comfortable revealing he could not read or write.

'I want people I'm familiar with helping me'

Melvyn, Mary and Graham had all repeatedly asked that Melvyn keep the same care worker. Due to the nature of his personal care it was important to Melvyn that he had a male carer. He previously had a male carer who he got on well with but that person left when the care provider changed. Melvyn's care workers changed frequently and each time this happened there were problems. In the space of two months he had three different care providers. The council re-let the care contract to a new provider.

Under the terms of the contract, care packages were cancelled every time the person was in hospital for 14 days or more. Of course, half the time Melvyn was in hospital because the system had failed to sort out his real problem. He usually stayed beyond 14 days because the system decreed that he was not 'safe for discharge', triggering a delay for assessment and subsequent care package during which he was transferred to a community hospital. The social worker no longer sorted this herself, instead placing the order via the council's new brokerage service.

Melvyn had a complex medication routine. Being unable to read he had always struggled to understand dosages and frequency. In 2005 Mary asked social services for help with medication management, particularly the nebuliser that he uses four times a day. After assessment he received domiciliary care three times a day, which was cancelled after three weeks with no recorded follow-up.

Melvyn left hospital (another preventable admission) on 27 October 2011. He had been in for more than 14 days and thus came home to new carers. The care agency flagged that the carers needed training in how to assist Melvyn with his nebuliser. The agency felt that his care plan was not specific enough, and it would not

risk doing anything not specified in it. In January, Mary said she was concerned that Melvyn was breathless and wheezy. In the same month, Melvyn had three review visits from the community matron and three visits from his case manager. Issues with the nebuliser were noted but no action ensued. The GP also visited to undertake a routine Frail & Elderly Annual Assessment. Melvyn couldn't understand why the GP asked him questions about his alcohol intake and advised him to contact the DVLA as he wasn't fit to drive. Melvyn couldn't drive!

More than 14 weeks later, the care workers were trained to monitor use of nebulisers. Melvyn was admitted to hospital with shortness of breath two days later.

The same thing happened in May when Melvyn was discharged from hospital after admission for gallstones. He then was given a different care provider. Issues with the nebuliser were unresolved and he again became short of breath. His care plan still said 'monitor use of nebuliser' despite the fact that Melvyn had never, as flagged by Mary in 2005, understood how to use it correctly.

Every time the care provider changed, a new risk assessment was conducted. In May 2012, the new care provider questioned whether 'doubling up' was needed to assist Melvyn in and out of his chair and bed. Until a risk assessment had been carried out, the care worker was not permitted to lift or move Melvyn. As a result he was 'cared for' in his chair for more than 24 hours until the new risk assessment had been completed. This occurred at a time when Melvyn was doubly incontinent, thought to be as a result of imbalance in his medications. The assessment concluded that doubling up was not required.

Melvyn had become angry and frustrated and often sent the carers away. He was labelled as difficult and non-cooperative.

'I want to be in control of my life and make my own decisions'

Melvyn liked to watch films, particularly at night, and wanted to decide for himself when to go to bed.

His last care call was at 9.00pm to ensure he took his medication, help him to bed, empty his catheter bag and attach the additional night bag. Melvyn repeatedly asked for a later call but under the new contract the supplier only undertook to attend within two hours of any preferred time. Other local agencies could have provided a later call, but they were barred by the terms of the contract. In any event the council's brokerage service did not understand why a later visit was important.

When the care worker arrived at 9pm, Melvyn often sent them away saying he wasn't ready to go to bed. Melvyn was seen as difficult and uncooperative. Sometimes carers would attach the extra catheter night bag and tubing while he was still in the chair, making it more difficult for him to go to bed later. Other times, Melvyn said he would do it himself but struggled to. As a consequence of overflowing bags and kinked tubes, his bedclothes were often wet in the morning and his infection problems worsened. Melvyn was frustrated and fed up so he started to use the Telecare service, late at night to get help. He learned that if he said he had fallen, someone (either an ambulance or a responder) would come promptly and help him to bed. However, no-one sorts out the catheter bags unless Melvyn remembers to ask them.

'I want to stay in my own home'

Over the years Mary and Graham have asked for help themselves as they struggle to support Melvyn. When Mary asked social services in 2011 for respite care, they were sent a leaflet on Age Concern. The last two years of managing endless appointments and a mountain of correspondence generated by 30 different

teams and professionals, attending endless assessments and reviews, being passed from pillar to post and being called out late at night all took their toll. The social worker was now concerned that they were both at breaking point. She and Mary felt it would be best for Melvyn to go into a care home.

Just 10 months before he was active and managing his various health problems reasonably well. His relationship with his friend gave him company and someone to go out with. He was adamant that he wanted to stay in his own home and dreaded the thought of going back into hospital or into a care home.

He was frustrated at the number of people who came to his home, often at the same time, from different agencies and who seemed more interested in their forms and paperwork than in him. When people did come to 'help' they just got on with their task and then left. He felt trapped in his own home and at the same time under increasing pressure from his sister-in-law and the social worker to consider moving into a residential home. The social worker made arrangements for Melvyn to visit a care home but he refused to get in the car when she arrived to take him there. Since then she has dropped off brochures for other homes for him to look through (although he can't read). He was made to feel selfish that he wasn't putting his brother first.

Most recently, his relationship with his friend ended. Melvyn said, 'She dumped me last week. She doesn't like me like this'. The system has changed how both she and the family see Melvyn - they now see him as an invalid.

Summary

Over the last 2 years Melvyn had spent 162 days in hospital of which, conservatively, 72 days (44 per cent) were avoidable. He had involvement from seven different agencies and 30 different teams and professionals. He went through 29 separate assessment processes. Given that the assessment process was repeated every time he re-presented or when one professional referred him to another, 66 per cent of these assessments were repeated. Overall Melvyn has had at least 74 assessments. He was most recently referred to the learning disability service whose assessment concluded that Melvyn did not have learning difficulties but simply had never learned to read or write.

He called Telecare 869 times, resulting in an ambulance attending 24 times and responders attending 61 times with follow up visits the next day in accordance with their 'best practice' policy. Telecare looked to withdraw the service because of Melvyn's 'abuse of the system'. He was not among their top 10 users.

Conservatively, over the last 12 months, the system spent £38,000 'helping' Melvyn without ever solving any of his problems or understanding what mattered to him. He is not alone - there are many more 'Melvyns' in the system. He was one of 2,145 people costing the local health care system at least £20,000 a year and in 2011/12 was the 236th most costly patient of 107,000.

He is now firmly on the glide path into the residential care system at a cost of around £250,000 over the next 10 years.

Jake

Jake's grandmother asked children's services for advice about Jake's difficult behaviour in January 1998. An initial assessment was carried out and the grandmother was given advice and guidance. After seven days, the case was closed.

Ten days later, a member of the same family contacted children's services anonymously to say that Jake was staying with his uncle and that they were concerned about his social skills and school attendance. The anonymous caller said they didn't think Jake's grandmother was coping with his care. Children's services sent a letter to the grandmother to see if she wanted to make an appointment with the duty social worker. Three days later, the case was closed.

Ten days later Jake's grandmother asked to see the social worker in response to the letter she was sent. She was given advice and the case was closed. Children's services carried out checks with the school to ask their opinion. No concerns were reported.

Three months later, another anonymous family member reported concerns about Jake's violent behaviour towards his grandmother. Children's services carried out checks with the school but no concerns were reported. Another letter was sent to the grandmother to ask her if she'd like to make an appointment with the duty social worker.

Two weeks later there was a meeting between the family and school. It was agreed that the school would continue to monitor the situation and Jake would attend a specialist centre that could give him support. Jake's family were advised to contact children's services if necessary for further advice.

The following month Jake's grandmother contacted children's services to say that Jake was not being looked after properly at his uncle's and that he should return to live with her. Children's services told the grandmother that because she has the parental responsibility, it was her decision. No further action was taken.

Jake's aunt contacted children's services a few days later to say that she was very concerned about Jake living with his grandmother. She was concerned that the huge changes Jake made since he had been living with her and his uncle would be undone. The social worker advised that because Jake was 15, he could live where he wanted and that his grandmother did not have parental responsibility.

A few days later the grandmother called children's services to say that Jake wasn't happy at his aunt and uncle's home and this was why he was living with her. The social worker advised her that she had parental responsibility and therefore it was up to her to decide, with Jake, what was right for him.

The following month, Jake's grandmother called yet again to express her worry about Jake's tendency to self harm and his deteriorating behaviour. She said she wasn't sure how much longer she could cope. The customer adviser from children's services advised her to contact her GP for an emergency appointment and to contact the school about a multi-agency support team (MAST) meeting.

Two weeks later children's services received a call to say that Jake was staying with a family friend after his grandmother had had a mental breakdown. Children's services prepared a referral and made plans to progress the case to an initial assessment. A child's plan was completed and the case was transferred to a local support team (LST) for a core assessment to be carried out.

After the case was transferred to the LST, a strategy meeting was held to discuss concerns about the grandmother's mental health. It was decided at the meeting that a follow up meeting was required because they didn't know enough about her mental health.

A couple of days later, Jake's grandmother called the out-of-hours service to say she was struggling to cope. She could not wait for the social worker to return to work after the bank holiday – she needed help immediately. The out-of-hours social worker passed it to the duty social worker to get back to Jake's grandmother after the bank holiday.

On the Tuesday, Section 47 enquiries were carried out and the core assessment was progressed to gauge risks to the child and to the grandmother. Two days later three contacts were received – one from the out-of-hours team in response to the call from the grandmother over the weekend and two from the police about a conversation with the uncle and a visit to his home.

Later that month, the Section 47 enquiries were completed and it was agreed that no further action was necessary but the core assessment would be continued.

Seven days later, on 4 June 2009, ongoing support was arranged through meetings and a child plan. Jake was moved into a specialist unit to assess his mental health and would stay there until the end of the month. Throughout June, the meetings and support were continued.

At the end of July when Jake was back home, the police contacted children's services to say that Jake's sister was ushered out of the house for teasing Jake about his self-harming. The social worker made contact with the family to talk about it.

A month later, Jake's grandmother took him to the police station to say she could no longer cope with his behaviour. Children's services provided advice and support and the grandmother took Jake home. Throughout September 2009 and January and March in 2010, more meetings were held and advice was given in response to contacts from the family via the out-of-hours service.

In July 2010, the case was closed because Jake moved into supported accommodation.

Part II

Recommendations: 'local by default'
a) What good looks like – principles and practice for effective service design

Effective service design means turning current scale assumptions on their head. Today's starting assumption is that since affordable capacity is unable to meet all demand, service must be limited, rationed by threshold or screened out, leaving remaining demand to be met by industrialised provision of standardised packages of service. As we have seen, however, it was the failure to help people to solve person-shaped problems that deepens the pool of failure demand which in turn threatens to swamp the system. The goal instead is to meet the need at the earliest transaction point and thus drain the pool to the real underlying levels that can be met by the redesigned system.

We are, to use an overworked phrase, at a turning point. Continuing down the present path leads to a collapse of all but the most critical public services. On the other hand, the case studies that follow demonstrate the existence of a once-for-all opportunity to unify and reconfigure services around a new common purpose – 'help me to solve my problem'. There is now a solid evidence base to show that services based on an understanding of individual need, context and views on what they require to live a good life – or in some cases to die a good death – can square the circle: that is, substantially improve the lives of those in need while at the same time draining costs out of the system.

Many of those providing the evidence have followed the Vanguard Method; others – mostly from the third sector – have arrived at similar results through a person-centred approach, applying good common sense in the assessment of needs and provision of services. Typically these services are locally-determined, responding to known local needs and historically unconstrained by the organisational features discussed in Part I.

The Vanguard Method was first developed in work with private-sector service organisations which have used it to improve service, lower operating costs and transform employee morale. In the public sector, the initial focus was on improving individual services (rent collection, benefits processing, care services and so on) in isolation. As those services became more effective, however, it became clear that behind the presenting need lay wider problems and issues that better delivery alone did not address. The first change of emphasis therefore was to redesign the service to take account of contextual issues – not just to provide faster benefits, for example, but to help claimants address the broader reasons for needing benefits in the first place. That led to a further discovery: the same people were commonly placing demands on a number of services. So the logical progression was to attempt to resolve the issues at family and community level, rather than individuals and services. This approach has come to be labelled 'locality' working, and it is a genuine breakthrough. As we shall show below, it has a profound impact on both the quality of individual lives and the quantity of demand for services, radically increasing the one as it reduces the other.

This kind of working is not a black box. Its essence is contained in the following simple principles. Effective services need to:

• Be 'local by default'

A thorough knowledge of the predictability of demand for services enables service providers to ensure that people who present as needing help can be met immediately by people with the requisite knowledge and skills to assess need and organise service provision. Real economies of flow replace imagined 'economies of scale'. Each locality is different; its needs can only be understood in a local context.

• Help people to help themselves

Services ask, 'What do you need to help you live a good life, or die a good death?' The focus is on strengths that allow people to make their own decisions rather than needs which render them more dependent on others and end up obliging them to lead the lives that others decide.

• Focus on purpose, not outcomes

Measures that relate to the purpose of the service from the users' point of view enable learning and improvement, as opposed to outcome-related measures that encourage cheating and hide failure demand.

• Manage value, not cost

Understanding demand from the customer's point of view, designing the service to absorb its variety (i.e. help people to solve their problems), and measuring achievement of purpose constitute managing value. The by-product of managing value is that costs fall out of the system. The by-product of managing cost is that costs go up.

Using these principles together has the key effect of reducing demand. Counterintuitively, mass-produced poor service is more costly than personalised service that meets individual need. By increasing service quality, and managing value and purpose, repeat demand that absorbs capacity and clogs the system is removed. What's more, there is a multiplier effect. The solving of for instance child A's problem (see page 26) effectively staunches failure demand leaching into the police, the justice system and schools from adult and children's care services.

Management's focus shifts from managing budgets and people to managing the system as a whole. Of crucial importance is management's knowledge of the predictability of demand; the maintenance of expertise to absorb that variety of demand; the ability for people at the first point of transaction to 'pull' on expertise required to assess or help people and measures which help them learn how improving these features drive costs down while services improve.

It is, in one sense, a shift from 'risk management' to 'knowledge management'. Part I showed how the current system's approach to risk management creates, in practice, high risk from a user's point of view as well as high cost to the public purse. It is also a shift from measuring for 'accountability' to measuring the things that will drive learning and improvement.

How it's done: housing services

As we noted, the learning and improvement process began with individual services. Housing benefits was one of the first Vanguard was involved with studying. Studying how their services met local demand, many local authorities concluded that official 'best practice' in the shape of service standards, front and back offices, activity management, and target-setting, was the cause of substantial sub-optimisation. Councils were conscientiously ticking the regulators' boxes, but their costs remained high while residents complained of poor service.

Redesigning services and matching upfront expertise to the variety of demand have since enabled them to provide much better service at lower cost. As examples, East Devon and Stroud councils reduced the time taken to process benefits to less than half the official target at a time when the number of claims was increasing. As services worked better to deal with people's requests 'right first time', failure demand dropped out of the system, freeing up capacity. This enabled East Devon to service 33 per cent more demand and Stroud 50 per cent more[11], in both cases using less resource.

Housing allocations and lettings services provided similar learning. Having analysed the nature of demand, Great Yarmouth Council realised that 50 per cent of people on the waiting list would never be housed (the system invited speculative applications and raised false hopes) and set out to explain to these applicants why and where they should seek more appropriate help. The 35 per cent with problems not primarily related to housing were visited personally to understand their needs in context, and then given help and support to resolve their real issues. This help and support was often practical rather than simply a referral to another agency. For example an elderly lady applied to move into a council property because she was concerned she could no longer look after her garden. The council arranged for help with tending to her garden, meaning she could continue living in her own property. The remaining 15 per cent who did need social housing were moved into appropriate lettings with support to resolve contextual issues. The result was a dramatic reduction in the waiting list and an improvement in first-time problem resolution from 30 per cent to 80 per cent, with no additional resource.

Like housing lists, rent arrears are another symptom of wider hidden issues. When it realised that central targets were focusing management's attention on collecting rent (the back end of the process), rather than resolving the issues that caused the arrears, a multifunction housing team at Bromsgrove & Redditch

11. Middleton, P (ed.) 2010 'Delivering Public Services that Work (Volume 1): Systems Thinking in the Public Sector Case Studies' Triarchy Press: Axminster

Councils decided to knock on the door of everyone in difficulty to understand their real need in context. As it progressed, it discovered (a common finding) that a large part of the problem was official policy rules that prohibited officials from doing the right thing. For example, tenants could not be offered alternative accommodation if they were in arrears, and Choice-Based Lettings 'points' determined 'need'. The new service design set these policies aside and focused on fixing underlying issues. When processed through the new design half the cases no longer presented as in need – the key measure of effective intervention. Housing service officials now see themselves as 'locality' officers, whatever their previous functional role. Rent arrears no longer operates as an enforcement or collection service but a means of getting residents' lives back on track.

Person-centred service: health and social care

The same story of official prescriptions distorting priorities and constraining initiatives is encountered in many TSOs. Community Lives Consortium (CLC) is a Welsh social enterprise which supports people with learning difficulties. Analysing demand, it found that structuring its work around the priorities effectively dictated by the commissioning authority's complex and bureaucratic assessment process was preventing it from listening and responding to its users' real needs. With better demand information, CLC could engage with its commissioners to develop more responsive approaches to user needs. Central to these were new jointly-developed highly personal service delivery plans based on their users' conception of 'a good life', some delivered in audio or video form rather than on paper. The more responsive way of working allowed CLC to eliminate substantial amounts of bureaucratic waste, much of the saving arising from replacing form-filling office work with activities that could be undertaken with service users as part of their support[12].

Learning is a key element in developing services that are truly person-centred; often the initial results are counterintuitive. A study of health and social care in Gloucester revealed that fully 86 per cent of demand presenting to local services was failure demand and that most staff time was spent on work that added waste rather than value. At a ratio of 75:25 per cent waste to value, it took 400 hours of work to deliver 100 hours of value. Meanwhile, just 5 per cent of the population consumed 50 per cent of the resource across the system.

Two small teams of nurses, social workers, physiotherapists and occupational therapists were set up to explore more effective ways of dealing with this demand, taking individuals that the local GP knew the system was having trouble coping with as test cases. It was quickly apparent that much of the problem was that for the NHS it was the 'wrong kind of demand'. While the NHS functions broadly as a 'fix-me' service for delivering one-off medical remedies for defined conditions, it has far more difficulty dealing with the more diffuse, often only partly medical 'help-me' problems that the troublesome cases were either partially or exclusively presenting with.

As the teams found, meeting 'help-me' demand requires skills and aptitudes that don't necessarily coincide with traditional professional boundaries. While technical and clinical competence remained important, for example for meeting 'fix-me' demands, it was paying attention to the social needs that made the biggest impact on better outcomes for both the individual and the system as a whole. That puts a corresponding premium on interpersonal, organisational and problem-solving skills as the key attributes needed for understanding and helping people to rebalance their lives.

12. Wilson R 2013 'Living the Life You Choose: The Introduction of the Vanguard Method into an Organisation Providing Support to People with Learning Disabilities' Systemic Practice and Action Research, January 2013

The service that has emerged from this learning is radically different from the previous model, being based on smooth end-to-end flow rather than repeated assessment and referral. It is also far simpler, demonstrating that the complexity and need for co-ordination of today's systems are a feature of system design and not an inherent part of demand. The ratio of value to waste work has moved to 80:20.

A similar social-care model, based on local teams linked to GP practices, has been developed in neighbouring Somerset. An initial study of 120 care users under the new system highlighted much reduced hospital admissions and lengths of stay, less need for social care support and avoidance of long-term placements in residential care homes. Reduced strain experienced by carers was another important outcome. A wider cohort study has since shown significantly better outcomes in terms of social care costs – effectively breaking the cycle of increasing dependency and decline.

b) Radical multi-disciplinary working that meets real need

All these cases illustrate the importance of understanding people and their families in context rather than limiting improvement work to the design of services. Since many people and families are presenting to many services with the same problems, the logical conclusion is to think of them all as forming part of the same locality-based 'help-me' service.

This is what happened at Bromsgrove & Redditch Councils. Having significantly improved the housing service, housing specialists have fanned out to work with other services to help all agencies understand the benefits of solving problems and reducing demand.

Stoke City Council went one step further, taking the radical decision to launch a comprehensive multi-agency initiative – across local authority, police, fire and rescue, NHS and TSO-provided services – to understand how people interact with the totality of public services: what did citizens need from public services in Stoke 'to live their lives well', and what did they get?

In the pilot area it was found that of 2,589 households, 5 per cent were placing demands on multiple services. The consequences were as described in Part I while capacity and resource was consumed in repeated assessing, referring and commissioning cycles, most underlying problems were left unsolved. Just as within individual services, the bulk of the issues were predictable across services, the highest-frequency issues being employment (67 per cent), managing finances (67 per cent), benefits and credits (42 per cent), suitability of housing (33 per cent), insecurity over housing status (29), distance from family and friends (25 per cent) and drug or alcohol dependency (25 per cent).

Working in multi-disciplinary groupings (i.e. across normal functional boundaries), the teams visited every family making demands on public services, with the aim of understanding the underlying problem and helping the person or family to find ways of solving it themselves. This confirmed the initial learning: the most important skills needed in the new-style public service are interpersonal – listening, interpreting and helping people to understand themselves. Specialist expertise is only brought in as needed and where proportionate to actual needs.

Based on geography not organisations, locality-based staff retain certain specialisms but place a greater emphasis on 'people skills' and prioritise relationship building. Multi-agency teams work together in individual neighbourhoods, come to understand local issues and get to know local families. These pioneering projects are breaking down barriers, improving outcomes and rebalancing the lives of customers to boost the economic and social wellbeing of whole communities. The results are profound. Citizens previously labelled as 'lost' are starting to live good lives, and demand across the spectrum of services is falling.
While cost savings as the consequence of providing better service cannot be predicted in advance, council chief executives label the size of the opportunity as 'staggering'. Predictions of financial savings in various local authority areas following two years of re-design run to hundreds of millions of pounds per annum.
If these figures can be proven, and if this pattern were replicated across the country as a whole, it would suggest that as much as £16 billion could be saved simply from a transition from a scale-based approach to an intelligent locality-working approach. The figures are based on estimation and extrapolation, but regardless of the precise amounts we are convinced that the scale of potential savings is enormous.

Could the public sector save £16 billion a year?

Could the public sector in England save £16 billion a year* whilst improving services? The answer suggested by multiple pioneer sites from across England is 'Yes'. Results from these sites demonstrate a staggering opportunity to improve outcomes for citizens and communities whilst reducing costs.

To put that into context, £16 billion is:

- A fifth of the UK Government's total public sector deficit (1)
- A sixth of the NHS Budget (2)
- Almost twice the total projected spend on care for older people in 2014 (3)
- Enough to fill the funding black hole facing local authorities projected by 2020 with £1.6 billion to spare (4)
- 500,000 extra nurses (5)
- 570,000 extra police officers (6)
- 380,000 extra hospital consultants (7)
- 570,000 extra social workers (8)
- 725,000 extra care assistants (9)
- 50 x the amount the Audit Commission say councils could save through "efficient assessment and review" (10)

*This is based on a simple extrapolation and an assumption of an average spend per capita, using 2001 population figures for target areas and England as a whole. The estimated potential for savings is based on knowledge gained through the direct, empirical study of several localities from across England. Each locality conducted detailed demand analysis to accurately reflect the nature and composition of their community, comparing this with broader socio-economic and demographic data to ensure that the localities selected were typical.

To estimate the savings for England, we have assumed that the estimated figure follows logically from known values. We are assuming that target areas are representative of the scale of opportunity elsewhere and having studied systems elsewhere we know that it is. However the figure arrived at extrapolates savings based on an assumption of an equal spend per capita profile across the country for the types of services profiled in this research, which will clearly not be the case. The only way to really know for certain how much this new approach could save is by trying it.

The £16 billion projection reflects savings in direct service costs, but does not include projected savings in overheads. If overheads also fall in line with reductions in demand and activity, the figure could be much greater still.

(1)– (10) References relating to the calculations provided above are in the Appendix.

c) An international example from the Netherlands: Buurtzorg

The experience of Buurtzorg, a Netherlands not-for-profit care home, shows that understanding demand in human terms and help for self-help are universal improvement principles.

Most home care in the Netherlands is based on the same scale-based 'production' assumptions as in the UK. Homecare was viewed as a product that could be provided more efficiently by dividing it into separate processes delivered by individual specialists according to strict specifications. However, gains in cost per hour were cancelled out by the need to coordinate and manage a complex fragmented process. The overhead drove demand for scale and yet more fragmentation in search of scale efficiencies, leading, according to Buurtzorg managing director Jos de Blok, to the hiring of staff 'who have an even lower level of training... [Some] have reached the level of administering pills and giving injections, others can do bandaging and some are allowed to do specialist tasks, such as connecting morphine drips. That is crippling for the motivation of the nurses and the quality of the care and, moreover, it costs society barrels of money.'

Buurtzorg has turned the process on its head. The focus of care provided by its (generalist) district nurses is explicitly the relationship with the client, the solving of problems and the rebuilding of patients' self-confidence as part of recovery. The organisation has shown that a single unhurried visit by a highly trained district nurse is more effective than several visits by specialised care workers each performing their allotted tasks – so care for a dementia sufferer, for example, might include sharing a chat and a coffee, feeding the cat, ordering medication, helping with bathing and dressing and even applying makeup. This way of working has increased 'unit cost' of interventions up to 30-40 per cent – but that is more than compensated by a 50 per cent reduction in total demand[13].

In the new regime, the role of district nurses has been revitalised. De Blok could see that the profession was dying a slow death in a care system that was 'suddenly all about production, protocols and administration. It was heading in the wrong direction.[14]' Now nurses offer complete care. This means that they may spend more time on basic tasks than previously, but since the job is now about relationships it is more varied than in the past. Nurses serve neighbourhoods of 10,000 people round the clock in self-managing teams of 10. Working with GPs, nurses see themselves as community builders, developing neighbourhood-level support for their clients from friends, family and volunteers. They use a weekly slot on local radio to advertise events and services, provide advice and put people in touch with one other.

Preliminary results show that Buurtzorg's patients consume just 40 per cent of the care that they are entitled to and half of the patients receive care for less than three months. As a result, patient satisfaction scores are 30 per cent above the national average and the number of costly episodes requiring unplanned interventions has dropped. The Buurtzorg approach to healthcare delivery has also led to reduced rates of absence through illness[15]. With no managers, communication lines are short, employees report greater work satisfaction, and in 2011 Buurtzorg was chosen as Dutch employer of the year.

13. As reported on the BBC Radio 4 'Today' programme, 27/05/13
14. Quoted in http://omahasystemmn.org/documents/2010-10-04ArtikelBuurtzorgInHetEngels.pdf (accessed 28/05/13)
15. Source: http://www.kpmg.com/global/en/issuesandinsights/articlespublications/value-walks/pages/netherlands.aspx (accessed 28/05/13)

Part III

Conclusions and policy implications

Conclusions

We reiterate here our four principles for achieving economies of flow in a locality.

Effective services should:

- **Be 'local by default'**
- **Help people to help themselves**
- **Focus on purpose, not outcomes**
- **Manage value, not cost**

This paper is a two-fold call to action

- In the first place, since the Vanguard-Locality approach makes no call on public funds – on the contrary, by freeing up capacity it liberates local resource – it offers encouragement for immediate action by service leaders to simultaneously save money and improve the lives of their citizens. Pioneers such as Stoke and Bromsgrove & Redditch councils have shown that it can be done.

- But in the longer term it also poses important questions for public policymakers and the role of regulation. This paper, and the research behind it, has been entirely self-funded by Locality and Vanguard. As such we acknowledge that it is limited in both scope and external scrutiny. We call on others – government, researchers, funders – to support efforts to investigate further the fiscal efficiencies and service benefits of adopting a 'local by default' approach. In particular, this should look to:

    - Replicate and review the assertion made in this paper about the extent (£16 billion) of potential public sector savings by a move to locality working.

    - Investigate the optimal delivery conditions for locality working, testing our assertion that in order to achieve efficiencies from this approach, 'local by default' principles may necessarily be required to extend not only to system design, but also to the very nature of delivery vehicles used. We would assert that remote, unembedded, and large scale bureaucracies (regardless of sector) will be inherently sub-optimal vehicles to deliver locality efficiencies.

Policy implications

Regulation

In the private sector and third sector, organisations are generally not constrained in the methods or philosophies they choose to work by – their management methods are part of the way they compete or deliver their social mission. Diverse thinking spawns innovation and experimentation, in time raising the bar

for all organisations. But this is not the case for public services, whoever delivers them. Official policy and the assumptions behind them obviously have broad influence over the extent and manner in which public services are delivered. However, in recent years, in an attempt to raise standards and decrease regional and local differences (the 'postcode lottery'), public guidance and regulation to enforce standards have become increasingly prescriptive, setting out not just what but how services should be delivered (call centres, front and back offices, shared services, central targets and service levels, payment by results, 'best practice'). Even where government departments have stated their aim is to reduce regulations and allow a provider's freedom to deliver e.g. the Department for Work and Pensions' 'Black Box' approach, new levels of bureaucracy and ultimately self-defeating outcome measures are introduced.

This is in no sense an argument for further privatisation, or to suggest that 'private good, public bad'. The point is that public sector organisations, and often those that deliver services on their behalf, are restricted by ubiquitous command and control, scale-based assumptions, in ways that other organisations are not.

This prescriptive approach to the regulation of public services has three main consequences.

- Regulation itself consumes more resources. As W. Edwards Deming put it, 'A regulation is justifiable if it offers more advantage than the economic waste that it entails'. Regulation is the fastest-growing and least accountable arm of the state. In 2005 the Better Regulation Task Force estimated that it could be costing the UK £100m a year – a huge burden of economic waste to set against any advantage created.

- By putting a straitjacket around method and work design, it stops innovation and experimentation in its tracks – another large hidden cost. It also halts learning and improvement.

- Even worse, if the prescriptions are wrong or based on faulty assumptions, they make performance worse – an even bigger concealed cost. This is the case in the public and increasingly the third sector, where regulation has locked in the 'diseconomies of scale' which make services unaffordable.

In other words, regulation is an inextricable part of the crisis in which public services now find themselves.

Purpose–measures–method

Bad regulation is part of the problem; good regulation should be an important part of changing public services for the better. Good regulation fosters innovation, challenge and learning. It places responsibility for improvement where it belongs, on the ground and on the front line. And it makes it the responsibility of service leaders to make choices about measures and methods for which they can be held to account.

It does this by focusing on purpose.

In any system, whether people are conscious of it or not, there is a systemic relationship between purpose (what we are here to do), measures (how we know how we are doing) and method (how we do it)[16].

16. See Seddon J (2008) 'Systems Thinking and the Public Sector' Triarchy: Axminster p82

The fundamental principle, followed by all the organisations to reach the results described in these pages, is that metrics are subservient but related to purpose – they should measure how well the organisation is responding to the customer's needs, from the customer's point of view. That in turn encourages experimentation with method – how can we respond better to customer need?

When regulators specify measures and methods, as in many of today's public services, the regulated focus is on compliance. Compliance becomes the de facto purpose, competing with and subverting the real one. Complying with measures (targets, service levels and PbR outcomes) and methods (assessments, treatment protocols) distorts priorities, causes people to cheat, and as we have seen, prevents services from meeting the needs of individuals. A large part of compliance is about avoidance of risk. But, as with cost, where management tries to avoid risk (for the organisation) it drives risk up (for the individual and/or community). Hence for example the recurrent tragedies in child care, the visible tip of a predictable iceberg. Organisations and people focused on form-filling, assessments and reporting to cover themselves (managing risk) had no time for their real purpose – paying attention to what was happening to the object of their care.

Like the organisations it regulates, regulation should focus on purpose, not outcomes, and the robustness of the relationship with measures and methods. To live good lives, people have different needs. Purpose needs to be developed in ways that help them solve problems, develop their independence and achieve the life goals that matter to them and their families. While regulators should articulate the purposes of service, the regulated are called on to make their own decisions about measures (how can we demonstrate that we are achieving the purpose for each and every client?) and methods (by what means are we going to help?), which regulation can test. The requirement for managers to make decisions about measures and methods ensures that learning and improvement take place where they should and will make clear to both providers and regulators when services are failing.

A focus on achievement of purpose

Creating an environment in which 'what works' flourishes will require fundamental changes to current approaches to commissioning.

As we have shown throughout this report, achieving better service with lower cost is a matter of applying a number of basic principles – 'local by default'; helping people to help themselves; focusing on purpose, not outcomes; managing value, not cost – with the aim of reducing demand as the main indicator of success.

For individual users of services, this means understanding their demand in context and providing the resources to help them meet it. This requires a different kind of commissioning. It cannot be achieved by services procured on the basis of standardised packages and price, and scale, as carried out today. Instead commissioners should want sound proof that service providers have the competence to understand demand in context and can show innovative capacity in provision of need. Recognising that these can only be effectively carried out locally ('local by default') they should beware of excluding competent local providers on grounds of size and consciously foster diversity in provision. Commissioning needs to encourage cooperation rather than competition among providers, and commissioners should look for awareness and willingness to embrace locality-type working. Every locality is different, and identifying the right providers for a particular purpose and place may take more time and effort up front. But the proof of the pudding will be that as needs are met, demand and costs will fall.

Payment by results

This is not the place to go into detail about why PbR fails in theory as well as practice. However, because of the increasing emphasis placed on it, and on outcomes-based management generally, across the spectrum of public services from the NHS to the Work Programme, it needs a mention here.

The most important consequence of effective service provision is a fall in demand. But precisely because it is a consequence – a by-product of more effective intervention – the fall cannot be specified in advance. It is a consequence of means, not ends. If it is specified in advance, it becomes a target and a de facto purpose which the organisation may very well meet, but at the cost of not doing other essential things or cheating. As the police well know, there is an easy and a hard way of meeting a target for crime-reduction: the easy way is to stop arresting people.

Moreover, because they do not distinguish between value and failure demand, current contracts offer providers no incentive to make demand reduction an outcome. To take a ubiquitous example, providers of call centres and back-offices are usually paid by volume, so the poorer the service and the higher the failure demand, the more they earn. Private-sector providers of custody services have an incentive to maintain volumes of people being processed through custody suites, often repeatedly, whether or not it leads to court proceedings. And outcome-based commissioning has traditionally led to providers focusing attention on the easiest to help, not the hardest ('creaming and parking'). As we have seen in Part I, the current system turns activities such as assessments and referrals into 'results' – sending a person on a parenting course or drug rehabilitation programme will be counted as an output for the organisation irrespective of the results for the individual, which are much harder to establish. Outsourcing on the basis of these kinds of targets (whether activity-based, output-based or outcome-based) simply locks in the enormous invisible cost of failure demand.

Like cost-based and risk-based management, outcome-based management does the opposite of what it promises. It fails the clients, demoralises staff and creates a cadre of management whose expertise is in compliance rather than improving achievement of purpose. Results – good outcomes – come from attention to purpose and means, not outcomes. And there's the paradox: managing by attention to purpose and means leads to improvement (better outcomes), managing by attention to output leads to distorted priorities, an unstable system and worse results.

Time to abandon a broken model

Public services have reached a parting of the ways. The issue is unfudgeable: the old scale-based model being the cause of today's crisis, it is impossible to graft the new one on top of it. The critical point, however, is that the crisis is reversible. We know how to help improve the lives of individuals and communities, and the good news is that it doesn't take more resources to do it. But it does take radically more effective use of existing ones.

Appendix

References for the figures used to put £16bn savings into context

1 http://www.ons.gov.uk/ons/rel/psa/eu-government-debt-and-deficit-returns/september-2013/stb--september-2013.html
In 2012/13 general government deficit (or net borrowing) was £82.1 billion, equivalent to 5.2% of gross domestic product (GDP) – down from 7.6% of GDP in 2011/12.

2 http://www.england.nhs.uk/allocations-2013-14/
Overall, NHS England has a budget of £95.6 billion to deliver the mandate.

3 http://www.nuffieldtrust.org.uk/talks/slideshows/projected-expenditure-care-older-people-2022
£9.07 billion was spent on social care for older people in 2011/12 to support 1.05 million people. See graph on slide 6 for 2014 projection.

4 http://www.local.gov.uk/media-releases/-/journal_content/56/10180/4053260/NEWS
Financial modelling by the LGA calculates that by 2020 funding cuts, coupled with rising demand for services, will create a funding shortfall of £14.4bn, with the widest gaps in funding falling on the most deprived areas of England, where demand for services is likely to be highest.

5 http://www.salarytrack.co.uk/average-nurse-salary
Based on an average salary of £32,000

6 http://www.salarytrack.co.uk/salary?kw=police+officer&lo=&type=permanent&currency=GBP&by=title
Based on an average salary of £28,000

7 http://www.salarytrack.co.uk/salary?kw=hospital+consultant&lo=&type=permanent&currency=GBP&by=title
Based on an average salary of £42,000

8 http://www.salarytrack.co.uk/salary?kw=social+worker&lo=&type=permanent&currency=GBP&by=title
Based on an average salary of £28,000

9 http://www.salarytrack.co.uk/salary?kw=care+assistant&lo=&type=permanent&currency=GBP&by=title
Based on an average salary of £22,000

10 http://www.careinfo.org/audit-commission-councils-could-save-300m-for-front-line-social-care-through-efficient-assessment-and-review/

Locality
33 Corsham Street
London
N1 6DR
0845 458 8336
www.locality.org.uk
info@locality.org.uk

in partnership with

VANGUARD

Locality is the trading name of Locality (UK) a company limited by guarantee, registered in England no. 2787912 and a registered charity no. 1036460.

Find us on Twitter, Facebook, Linked In

locality

VANGUARD