DEVELOPING THE INDUSTRIAL STRATEGY CHALLENGE FUND HEALTHY AGEING (ISCFHA): a technologically enabled ecosystem for healthy ageing.

Sarah Harper and Kenneth Howse
Oxford Institute of Population Ageing
University of Oxford
July 2019

www.ageing.ox.ac.uk

Table of Contents

Preface

The Oxford Institute of Population Ageing, University of Oxford was commissioned by UKRI to produce a review of previous government funded initiatives in healthy ageing, to inform the development of an approach to the Healthy Ageing Challenge by providing an independent view of relevant learning from previous initiatives in the field of healthy ageing.

All the case studies we selected have specific objectives which link with healthy ageing in one way or another. While the focus is on technology based demonstrators relevant to the promotion of healthy ageing, and the issues, problems and challenges that arise in designing, implementing and evaluating them, a small number of healthy ageing programmes that do not have a service or social and behavioural focus are also included. Some of them set out to promote increased personal engagement with health among the general population or specific sub-groups within it. Others aim at service improvement or innovation with a focus on preventive action for older people who already have health problems or disabilities. In both kinds of programme, technology is seen as a potential enabler of change. In a class of its own, there is a public health programme which focuses on the social and physical environment as an enabler of lifestyle change for obesity reduction. The significance of this particular programme, despite the absence of a technology focus, is an example of an attempt to construct a 'whole system' approach to a major public health problem.

We conclude that recent ventures around healthy ageing have been limited in scope, developed in isolation of one another and rarely reached significant scale. There are also significant gaps in the current investment ecosystem, across the development pipeline from early-stage development through to scaling, and there is a need for much greater collaboration across and between different stages and sources of funding.

To explore some of the more complex interdependencies within and across themes/ sectors, there is value in investing in some place-based responses which can provide opportunities to explore and evidence how a local ecosystem of public and private organisations can support the development and scaling of solutions. Development of partnerships and mechanisms and frameworks for collaboration will be an essential, and sufficient resource needs to be assigned to establishing these.

The approach suggested, which builds on the original work by the Institute in contributing to the original specification for the Healthy Ageing Challenge, is ***to decrease the probability of transitioning from one dependency state to another and ameliorate the impact of that transition*** and thereby attain an **extra 5 years of active healthy life**.

Sarah Harper
Kenneth Howse
Oxford July 2019

Executive Summary

1. Designing, implementing and evaluating a demonstrator style project for the ISCFHA.

A key challenge for the UK is **to decrease the probability of transitioning from one dependency state to another as we age, to ameliorate the impact of that transition as and when it occurs,** and to thereby **to attain an extra 5 years of active healthy life.**

The practical challenge is to deliver products and services which demonstrably support older people to remain active, productive and independent, thus delaying transitions into more intensive care services. A strategic response to this challenge requires two pillars:

1. The first involves interventions to promote positive behaviours throughout life, to maintain health and delay the onset of care needs.

2. The second is about creating inclusive products and services which support older people, and their carers, to maintain their wellbeing and independence.

We argue that a Challenge Programme which focuses on the potential of new technologies to contribute to the promotion of healthy ageing, needs to

- be explicit about the modalities by which change is to be effected
- emphasise the importance of novel framing of challenge
- place people at the centre not technology

We identified a series of lessons

1. Vision: clarity about the purposes being served by running a demonstrator entails clarity about the kind of knowledge or know-how that is expected from the evaluation.

Ensure all key players understand and share a sense of the vision. The challenge is how to get an intervention or innovation 'up and running' or embedded in routine practice in a way which can be expected to produce the desired outcome at a population level. The basic decision to consider in planning a demonstration programme is whether it is necessary or desirable to run an experiment, and if not, to be clear about the purpose of the demonstrator and the nature of the evaluation, including the appropriate criteria of programme success and failure.

2. Identify measures of success

Demonstration programmes can fail in different ways because they have different purposes in view. Experimental demonstrators can fail to produce robust evidence of positive impact or it may prove impossible to implement the original experimental design.

- The *Whole* System *Demonstrator* and *United4Health* ran into both these problems. A programme that has implementation at scale as its primary aim fails if the innovation is not up and running by the end of evaluation period or if participation targets have not been met.
- *HealthSpace* and *Dallas* failed to the extent that the levels of public participation they actually achieved fell way short of their ambitions or expectations. They failed in terms of their primary objective.
- *Healthy Towns Programme*, which set out to test 'whole systems' approaches to obesity reduction, and in the event succeeded in implementing a whole system approach in only one of its 9 demonstration sites also fell short.

3. Incorporate specific targets: for longer-term goals like spread, sustainability and scale-up

Measures of success need to be integrated within the project design:

- Scale-up – within the setting
- Spread – transfer to new settings
- Sustainability – being maintained long term, adapting as required

Reasonable scale-up for a given condition will drive forward population health benefits at scale but 3 years is too short.

4. Recognise Complexity: avoid assumption that the issues to be addressed are either simple or *complicated* (hence knowable predictable and controllable) rather than *complex*.

Assess the nature and extent of complexity in the programme and ensure that emergent and adaptive measures are used to address these issues

- *complexity* in multiple domains poses a massive challenge to scale-up, spread and sustainability.
- programmes which were *complicated* in multiple domains proved difficult, slow and expensive to implement.
- programmes which were *complex* in multiple domains did not become mainstreamed - or if mainstreamed - did not deliver key intended outputs.

A recurrent theme for all the technology demonstrators is that planners seriously under-estimated the complexity of the systems - the socio-technical networks - within which a newly introduced technology has to be made to work. Failing to account for complexity is one of the main reasons why attempts to implement at scale so often fail to keep to programme timetables or achieve participation targets. The recommendation of the HMHM programme evaluation, that a formal analysis of complexity needs to be built into planning process for future digital health programmes, deserves serious consideration.

5. Consider "technology readiness": define the circumstances in which it would be useful to conduct an assessment of 'technology readiness' as part of the planning process

Whether or not technology demonstrators pose special problems or add an extra layer of complexity to an already complex situation it is hard to say. Lack of 'technology readiness' in the socio-technical network being targeted for implementation has been proposed as a possible explanation for programme failure.

6. Trade-off: acknowledge there is a trade-off between programme ambition and scale, on the one hand, and a manageable timetable, on the other.

The challenge of complexity is a challenge for timescales. Evaluations are constrained by the time periods they cover, and it is possible for programmes to be judged prematurely as failures - because the initial timeframe for achieving particular targets was unrealistic. Inflexible timetables may not allow time to respond to emergent problems. The reports we saw agreed in thinking that **3 years is not enough time for the evaluation of a large and ambitious programme. Our recommendation is a 5 year plan.**

7. Integration: findings from separate evaluations of distinct interventions need to be integrated so as to provide an easement of overall programme impact

Control over the quality of an intervention, fidelity to an agreed model, and the conditions of delivery are important for an experiment. The decision to loosen that control and allow for multiple interventions with different target populations can create difficulties for evaluation.

8. Leadership: directives for both ownership and local leadership to be built into initial plan

It is important to consider how local ownership may be built into a national programme with overarching objectives. In addition, it is necessary to establish overall leadership since complex programmes often suffer from outsourcing of control and coordination. The case studies illustrate different approaches to the balance between national objectives and local implementation. The recent history of technology demonstrators suggests a shift towards different forms of local 'ownership' for the implementation of national programmes.

9. Talent pipe-line

A talent pipe-line is essential to retain knowledge and expertise in the workforce developed over the course of the programme.

10 Adapt: build in adaptation as the programme evolves.

Collect and reflect on emerging data, be creative, but control over ambitious growth, since projects that evolve organically are vulnerable to over-ambitious extension and scope creep.

2. Developing the industrial strategy challenge fund healthy ageing programme: joint lessons with innovation, investment and funding scope[a]

Recent ventures around healthy ageing have been limited in scope, developed in isolation of one another and rarely reached significant scale. There are also significant gaps in the current investment ecosystem, across the development pipeline from early-stage development through to scaling, and there is a need for much greater collaboration across and between different stages and sources of funding.

To explore some of the more complex interdependencies within and across themes/ sectors, there is value in investing in some place-based responses which can provide opportunities to explore and evidence how a local ecosystem of public and private organisations can support the development and scaling of solutions. Development of partnerships and mechanisms and frameworks for collaboration will be an essential, and sufficient resource needs to be assigned to establishing these.

We thus suggest a programme of linked eco-systems

1. Creating a productive, effective innovation ecosystem

- Supporting partnerships between researchers from the social, life and medical sciences, technology, creative arts design and humanities, innovators, industries and businesses to support more effective knowledge exchange and development.
- Supporting navigation through the various funding sources and mechanisms at different stages of the development and investment pipeline.
- Improving the quality and effectiveness of current early-stage innovation activity, steering this towards prevention, inclusivity and evidence-based, user-informed design.

2. Creating an viable, attractive funding ecosystem

- Providing sources of early-stage development funding which supports higher-financial risk, high social impact return ventures to develop; socially-minded ventures to scale sufficiently to be 'investment-ready'; and for alternative funding models, which enable high social impact, slower or lower-return ventures to scale and achieve commercial viability.
- Incentivising the entrance of key industry players such as large retailers, manufacturers, financial services and creating channels for new and more established ventures to reach the mainstream market.

[a] The Institute was asked by UKRI to produce a short summary linking the main lessons from our report and those emerging from *Healthy Ageing innovation ecosystem and funding and investment landscape* (Centre for Ageing Better July 2019) which covered the investment challenges and opportunities

3. Creating an ecosystem for healthy ageing enabled by technology

This would involve a multidisciplinary, multi-sector team working together to explore opportunities for the

- **Co-creation of solutions** by bringing together technology push and market-pull factors across a diverse range of viewpoints, constraints and knowledge levels so as to sustain the exploration of new scenarios, concepts and related potential products and solutions.
- **Exploration** by engaging all stakeholders, especially user communities, at an earlier stage of the co-creation process to discover emerging scenarios, usages, commercial models and behaviours through live scenarios in real or virtual environments.
- **Experimentation** that implements appropriate technological artefacts in vivo to generate live scenarios involving large number of users whilst, in parallel, collecting data for analysis.
- **Evaluation** that assesses innovative concepts as well as related technological artefacts in real life situations through various dimensions such as socio-ergonomic, socio-cognitive and socio-economic aspects. It is the observation of what happens when technology confronts user value models that roots a Living Lab in experiential rather than experimental learning.

Process

- Identify **contrasting places** with contrasting demographics, economies, social groups which would provide the real-life environment where systems, services and processes relating to healthy ageing could be developed at speed, evaluated, de-risked and integrated with the needs and ambitions of the local community.
- Pipelines of talent and innovation into, across and as outcomes of the programme is essential. Integrate a strong **pipe-line** of products and services drawn from innovators, developers, and businesses across the UK.
- Define the balance between **national objectives and local implementation**. Adaptation to local conditions is increasingly seen as a precondition of success.
- Manage the tension between **innovation and implementation**, especially when continuing evolution of the technology adds to complexity, needs to be recognised and integrated into the original design

Framework

- Produce a clear vision outlining the purpose, the kind of knowledge, and the expertise which is expected
- Identify measures of success whether the implementation of the experimental design or of the outcomes of the experiment.
- Incorporate specific targets within the project design: scale-up within the setting; spread and transfer to new settings; sustainability over the long term.
- Recognise complexity avoiding assumption that issues are either simple or *complicated* – knowable, predictable and controllable - rather than *complex*.
- Define ambition and scale with recognition that single time limited programmes cannot address complex system problems without time.
- Leadership from the Challenge Team for the whole programme and local leadership for large and small projects within is essential
- Build in adaptation as the programme evolves.

Part 1

1.Introduction and Summary

The Healthy Ageing Challenge:
To deliver new products and services which demonstrably support older people in the UK to remain active, productive and independent, delaying transitions into more intensive care services

The Oxford Institute of Population Ageing, University of Oxford was commissioned by UKRI to produce a review of previous government funded initiatives in healthy ageing, to inform the development of an approach to the Healthy Ageing Challenge by providing an independent view of relevant learning from previous initiatives in the field of healthy ageing.

The practical challenge is to deliver new products and services which demonstrably support older people in the UK to remain active, productive and independent, delaying transitions into more intensive care services. A strategic response to this challenge requires two pillars:

3. The first involves interventions to promote positive behaviours throughout life, to maintain health and delay the onset of care needs.
4. The second is about creating inclusive products and services which support older people, and their carers, to maintain their wellbeing and independence.

While early life interventions are generally beyond the current initiative, understanding of programmes around mid and later life - on work, design and nutrition for example – are important. In addition, older people are not a homogenous group. Their needs and aspirations are individual and complex. Nevertheless, innovations can support independence and slow the transition from one level of dependency to the next. In particular we can identify transitions from active adulthood to a degree of physical disability to declining cognitive capacity, particularly dementia, to with multiple-morbidities and high acuity needs. The approach suggested which builds on the original work by the Institute in contributing to the original specification for the Healthy Ageing Challenge is ***to decrease the probability of transitioning from one dependency state to another and ameliorate the impact of that transition*** and there by attain the extra 5 years of active healthy life.

The ISCFHA Challenge is a £98M investment over 3 years aimed at supporting innovation and developing new offers for health ageing. It will support delivery of the Ageing Grand Challenge, investing in an area where the UK is well placed to exploit growing domestic and international markets through the development of innovative, data and tech based, products and services. Within the proposed approach to the fund is a £75m programme comparing, validating and improving products; developing new business models to combine multiple innovations into new services; and demonstrate their use at scale.

As the original bid described, specific technology sectors relevant to the care system are projected to grow rapidly. The market for medical and clinical wearables is expected to grow to $8.3bn in 2020. The Health and Care IT sector is expected to grow by 16% annually to $203bn in 2025. Estimates suggest the smart home market will to grow $56bn by 2020, and assisted living to $3.96bn by 2020. The market for care assistance and automation robots market is expected to grow rapidly too.

The original scope suggested a small number (2-4) of place-based development and demonstrator facilities for testing and validating new products and services. Each centre could focus on preventing transition from one level of dependence to the next, or specifically on innovations to manage physical disabilities, cognitive disabilities, or modifying the home. The Challenge also conducted an extensive engagement with industry. This has indicated that there is widespread interest in the sector. However, analysis of the level of commitment from industry in terms of scale and co-investment suggests that the proposed scale and scope of the demonstrators was unlikely to be feasible.

The review summarises key learning relevant to the design of the ISCF Healthy Ageing Challenge programme, thus providing an independent view of lessons from previous programmes to inform the design of the Health Ageing Challenge. The review focuses on government-initiated 'demonstration programmes' or 'pilot programmes' which have a strong technology or service focus and are of relevance to the promotion of healthy ageing. It primarily highlights the issues, problems and challenges that arise in implementing and evaluating programmes that involve the use or adoption of innovative technology and or services at scale in settings and contexts relevant to the promotion of healthy ageing. In addition, it considers the important social and behavioural element of this challenge – which plays a more significant role than in some of the other Challenges.

The review takes the following approach.

- A review of past government funded initiatives in healthy ageing
- Presentation of lessons learnt relevant to the current ISCFHA initiative
- Designing, implementing and evaluating a demonstrator project.

2. Aims

The aim of this report is to review government-initiated UK 'demonstration programmes' or 'pilot programmes' which have both a strong technology focus and have some relevance to the promotion of healthy ageing. We are interested primarily in the issues, problems and challenges that arise in *designing*, *implementing* and *evaluating* programmes that involve the use or adoption of innovative technology *at scale* in settings and contexts relevant to the promotion of healthy ageing.

We take a case study approach to the analysis of the conditions of success and failure in a selection of demonstrator programmes. The analysis is based on published evaluation reports, and our judgments are thus based on a synthesis of the conclusions drawn in the programme evaluations.

We have also incorporated some of the reflections and ideas laid out in a small number of theoretical/analytical papers by independent researchers who have been involved in the evaluation of large-scale UK programmes.

We have not considered published evaluation of trials of health interventions that would not count as demonstration pilots, and we excluded the very large number of small local initiatives, initiated not by central government but by PCTs or CCGs or AHSNs or LAs.

3. Considerations

The relationship of science to technology

The literature identifies four genres of technology research. These may be viewed as generations of research which develop and were built upon, or which ran concurrently with each other within the one project or programme.

- Technical design to provide proof of concept
- Experimentation, largely RCTs
- Qualitative studies to understand user - client and carer – experience
- Institutional exploration to elucidate competing interests.

Many of the previous programmes in assisted living technology were hampered due to a misunderstanding of the role and complexities within the overall system.

Tension between science/academic research and technology

This reflects science and technology having different purposes, with

- science aiming to produce models and theories that explain nature and (sometimes) allow predictions of its behaviour
- technology is developed to generate useful artefacts.

Technology focus on what works reliably rather than what is new, and because the costs of failure are typically high, they normally rely on a legal process of accreditation, based onan established and formalised body of knowledge, to ensure the quality of their work. Science, on the other hand, the ability to explain natural phenomenon requires the phenomena to be analytically tractable. This typically requires a degree of simplification to reduce the number of variables and interactions under consideration **(Pavitt, 1999).**

Problems versus process – process is important

Technologies as *things* that solve problems v. *innovations* that transform process
How the technology is *expected* to behave (artefact) v how it *will* behave (a complex of artefacts and knowledge). Technologies are comprised of artefacts, that generate artificial functions, techniques and the wider institutional regime required for them to operate.

> ***"Consequently, technical change isn't just a one way process of changing the world to match an idea, or something determined by static social structures. Insteadit is generated by a distributed, often contested, co-evolutionary process that involves incremental improvements and radically new combinations, in which understanding and artefacts both change in a complex combination of deliberate design and unintended outcomes". (Pavitt, 1999)***

Lock-in[1]

There needs to be a recognition that technologies may lock into path dependent trajectories, particularly if change is subject to

- increasing returns to scale where the more it is used the lower the costs become
- learning effects where the more we use something the better we collectively perform
- network effects where information and physical networks become more valuable as they grow in size
- changing expectations as increasing adoption reduces uncertainties about quality,
- performance and permanence in the market

All these phenomena generate positive feedback loops and strong selection effects that reinforce the use of a particular technology.

Misunderstanding technology as "gadgets"

This way of thinking of technologies as "gadgets" misses

- the cognitive element,
- the imposed nature of technological functions,
- the tacit knowledge needed to develop and operate a technology,
- the dependence of technologies on wider connections to complementary devices, institutions and forms of knowledge.

There needs to be recognition that technologies may generate profound and irreversible changes in commercial, social and physical environments.

Part 2

4. The Case Study Approach

The review takes a case study approach. Case studies fall into three types:

- the intrinsic - examined for its own sake
- the instrumental – a selection of cases to examine patterns or lessons
- the collective – the collection of specific data from different cases to enable generalisation

The advantages of the case study approach include the examination of data within its own context. This allows an exploration of micro and macro data and processes, which enables an understanding of issues in real-life situations, including gathering information concerning strategies used, the reasons for strategy use, and how the strategies are used in relation to other strategies.

Government-led demonstration programmes have taken many forms in the UK. They do not all have the same purpose and they may not actually be called demonstration programmes or demonstrator projects. A recent paper by Ettelt *et al*[2] usefully distinguishes between *piloting*[b] for experimentation, early implementation ('pioneer') demonstration ('demonstrator') learning. With the possible exception of early implementation, what these various purposes have in common is an evidence-based approach to the introduction of innovation in policy or practice, typically including some evaluation. Demonstrators or learning pilots typically also produce some kind of toolkit or guidance. Furthermore, randomised controlled trials (RCTs) may be reasonably described as policy pilots under certain conditions. For example, if they can be characterised as complex interventions which require cooperation from different 'stakeholders' or 'users' or 'adopters' *and* investigate problems of widespread implementation or scaling-up.

Recent UK government practice gives us examples of demonstration programmes which

- emphasise technology
- focus on the organization and delivery of services
- take a public health approach

[b] Experimentation, at least in theory, assumes genuine uncertainty about the superiority of what is being tested over the status quo. An early implementation pilot is a first step towards national roll-out – aka 'phased implementation'. Demonstrators, in the rather restricted sense of the term used by Ettelt et al, select the 'most promising localities to show others how to implement successfully'; and learning pilots ('trailblazers') focus on ways and means of overcoming implementation challenges and improving processes and outcomes. Since pilot programmes may be set up for multiple purposes, we should not assume that all pilot programmes can be slotted into one and only one of these categories.

The case studies we have included are presented in Part 3and our commentary is arranged under four main headings to reflect common problems across all these demonstrators irrespective of their specific goals and objectives.

1. The rationale for running a demonstration programme - what it is trying to achieve. There is of course a basic difference in this respect between demonstrators with and without a technology focus. Broadly speaking, all the technology demonstrators are concerned with realising the promise of new technology for health and social care services and for population health improvement more generally. They focus on the potential of new technologies as enablers of innovation in services.
2. Problems of recruitment and public participation or adoption.
3. Common challenges of implementation in innovations which require change in complex systems.
4. Issues of evaluation, this too relates to complex systems

UK demonstrator programmes/policy pilots reviewed with a strong technology focus
HealthSpace (2007- 2013)
Whole System Demonstrator (2008-2011)
WSD Action Network (2008-2011)
dallas (2012-2015)
Home and Mobile Health Monitoring (2015-2018)
United4Health (2013-2016)
NHS Test Beds Programme (2016-)
Social Care Digital Innovation Programme (2017-)

UK demonstrator programmes/policy pilots reviewed with a service/public health focus
Healthy Community Challenge Fund/Healthy Towns
Partnerships for Older People Projects (POPP)
NHS Diabetes Prevention Programme

5. Purposes and Objectives of the Demonstration Programmes

Why run a demonstration programme?

Government-led demonstration programmes have come in many shapes and sizes in the UK.

Experimentation, at least in theory, assumes genuine uncertainty about the superiority of what is being tested over the *status quo*. This is the state of 'equipoise' which is usually required for a randomised controlled trial (RCT). In practice, a policy experiment can fail to meet this condition and nevertheless be quite properly described as an experiment. It might be gathering evidence to reduce uncertainty about 'what works' or to strengthen the case for government action (and public expenditure). Either way the fact that the programme is regarded as an experiment presupposes that it is an open question whether or not the innovation being tested can actually deliver the expected benefits. Whether or not we think that the *Whole System Demonstrator* met the condition of equipoise, it was judged that a methodologically rigorous experiment - in this case an RCT - was needed in order to generate 'credible evidence' of the value of an innovation in clinical practice.

Policy experiments typically require sustained cooperation not only from multiple 'stakeholders' in one or more provider organization, but also from members of the public who use the services. Public health initiatives which aim at bringing about behaviour change in the general public offer an example of a somewhat different kind of experiment. In such cases, what is driving organizational change is not so much the hope of clear benefits for 'service users' as the hope of higher levels of engagement among the general public with collective goals - like keeping healthy.

An early implementation pilot is a first step towards national roll-out – *aka* 'phased implementation'. Demonstrators, in the rather restricted sense of the term used by Ettelt *et al*, select the 'most promising localities to show others how to implement successfully'; and learning pilots or 'trailblazers' focus on ways and means of overcoming implementation challenges and improving processes and outcomes. The distinction between policy experiments and other kinds of policy pilot raises an obvious question: why pilot an initiative without running it as an experiment?

With the possible exception of early implementation, what policy experiments and other kinds of policy pilot have in common is an evidence-based approach to the introduction of innovation in policy or practice, which means a need for some kind of evaluation, ideally independent and transparent. The experiments test whether or not an intervention or an innovation works (and clearly some experiments are better designed than others). The experiment is itself the evaluation, and a well-designed experiment will have a clearly specified question that it is designed to answer.

As for the demonstrators or learning pilots, the evaluation typically focuses on the practical problem of how best to implement or effect changes in practice – how to get an intervention or innovation 'up and running' or embedded in routine practice in a way which can be expected to produce the desired outcome at a population level. The problem is often presented in terms of 'scaling-up' - of "moving from a local project to one that is 'business as usual'"[c], or from ideal conditions to real-

[c] Greenhalgh ref

world conditions. The rationale for piloting an initiative without running it as an experiment is that the *primary* objective is to find a way of getting something 'up and running' and embedded in routine practice.

The evaluation report for the HMHM programme makes use of a 3-fold distinction between goals taken from Trish Greenhalgh in its assessment of progress over a 3 year period.

- **Scale up involves moving from a local project to one that is 'business as usual'.**
- **To spread an innovation is to transfer it to new settings.**
- **Sustainability is demonstrated if an innovation is maintained long term and adapted as necessary.**

Many of the programmes included in this report as case studies are quite explicit that the implementation of innovation at scale is a core programme objective. Dallas was intended to *Deliver Assisted Living Lifestyles* ***at Scale***. The evaluation report for the United4Health project concludes that the "deployment sites have shown that it is possible to deploy telehealth solutions **at scale**" by which they mean that they succeeded in redesigning care processes so as to integrate telehealth into routine care delivery services. In other instances, the programmes are clearly aiming at the diffusion of an innovation through the population (*HealthSpace*) or its integration into routine practice for service delivery (*Whole System Demonstrator* or the *Combined Predictive Model*)[d]. The same applies to some of the programmes which do not a have a specific technology focus. The NHS Diabetes Prevention Programme (1st wave) is intended "to deliver **at scale**, evidence based behavioural interventions for individuals identified as being at high risk of developing Type 2 diabetes". *Home and Mobile Health Monitoring* was a 3-year programme which aimed to "move from small scale pilots to large scale deployments informing **scale-up and spread**". This is the only case study which explicitly distinguishes between scale-up and spread. See Box 1.

This means that failure for a demonstrator or a learning pilot is quite different from failure for a policy experiment. A well-designed experiment may fail to produce robust evidence of positive impact, which is one of the ways in which the *Whole System Demonstrator* failed[e]. Experiments may fail of course in other ways, for example, by being poorly designed (e.g. inappropriate methods) or by a failure properly to implement the design (e.g. recruiting the sample as specified in the design).

A demonstrator, on the other hand, fails if the innovation is *not* properly incorporated into routine practice ('up and running') by the end of the evaluation period; it fails if it does not 'deliver at scale' an innovative service that worked as expected. This was how *HealthSpace* failed[f]. In some of the other programmes, we can use other declared programme objectives as criteria of success or failure. The *Healthy Towns Programme*, for example, was intended to 'stimulate' the adoption of innovative

[d] Much of the work undertaken in AHSNs is concerned with the adoption and spread of innovations across the country, as with the Electronic Frailty Index.
[e] Henderson C *et al* 2013
[f] How this particular criterion of success or failure is to be operationalised is a different matter. In some cases, as with *HealthSpace*, it may be blindingly obvious that it failed; in others, perhaps less so. By the same criterion - achievement compared with ambition - *dallas* also failed.

approaches to tackling a major public health problem (obesity), and it was the idea of a 'whole systems' approach that supplied the evaluators with a criterion for success. What Government hoped to see were several variations on a single theme - a 'whole towns' approach. In the event, only one of the nine test sites succeeded in putting together a programme of interventions that qualified as a 'whole systems' approach. If one of the aims of *dallas* was to provide a stimulus to digital health industries, this would provide evaluators with a criterion for success or failure (though none of the published evaluations appears to assess the programme from this point of view).

The WSD had two main (incompatable?) objectives:

- **to investigate the effectiveness of remote care technologies with a view to producing the kind of evidence that would persuade clinicians to incorporate them routinely in their practice**
- **to drive whole systems redesign supported by advanced assistive technologies, i.e. show how telehealth (TH) and telecare (TC) can be used to move towards a more integrated approach to providing managing people with health and social care needs in the community**

One of the main rationales for the WSD was that previous pilots or trials of remote care technologies had been too small or insufficiently rigorous to provide robust evidence of the evidence of the effectiveness of the TH and TC technologies. The plan therefore was to run a trial with sufficient power to answer the questions about impact (i.e. an experiment). At the same time it was recognised that the health and social care system could be much improved if services were more integrated (i.e. a demonstrator).

The evaluators took the view that one of the most fundamental problems with the design of the WSD was the tension between adherence to the requirements of an RCT and the challenges of implementing a new technology that required different groups of people to adopt new ways of doing things, i.e. changes in routines and behaviours and structures*. The RCT design got in the way of running a 'demonstrator' pilot by effectively blocking opportunities for (i) adaptation to the particularities of local contexts, and (ii) adaptive learning and the ability to respond flexibly to emergent challenges. This tension was manifested in an initial divergent understanding of the programme objectives among key stakeholders. Local managers had a rather different view of what they were doing (getting additional resources to promote local change) than the DH (running a policy experiment that combined a RCT and whole system redesign), and one consequence of this ambiguity was disappointment among the local managers when it was eventually clarified.

There is a further difference between experimental and non-experimental pilots. Unless a pilot is designed as an experiment which requires 'intervention fidelity' - keeping to an agreed template for design and delivery - there is no good reason not to change an implementation strategy in the middle of a programme. There is nothing wrong with adaptive learning in a learning pilot - which may include changing modalities of delivery - *if* the primary objective is to get an innovation 'up and running' at scale. A trial and error approach to implementation is warranted within the programme. If it is desirable to allow (or even encourage) adaptive learning in a demonstration programme, it may be a mistake to run it as an experiment. Indeed this seems to have been the reason why the

Whole System Demonstrator failed not only as an experiment but also as an attempt to bring about organizational change.

Since pilot programmes may be set up for multiple purposes, it cannot be assumed however, that all pilot programmes can be slotted into one and only one of these categories. It is in fact quite clear that some of the programmes included in this report combine a policy experiment with a demonstrator or learning pilot. There are indeed only three programmes in our case studies which can be unambiguously classified as policy experiments, namely the *Whole System Demonstrator*, *POPP*[g], and *United4Health.* The description fits not only because they are explicit about their objectives - they say that they are designed to produce evidence of the impact of remote care technologies on measurable outcomes – but also because they include control groups, and to that extent at least have consciously attempted to conform to established standards for what constitutes a well-designed experiment in this field. It is interesting to note that the *Summary Care Record Programme*[h] and the *Healthy Towns Programme* seem to have been originally conceived as experiments, in the event they were not implemented and run as such[i].

Little is achieved if we try and blur (or challenge) the distinction between experiment and demonstrator or learning pilot. If a demonstrator is set up in order to test an organizational approach to the implementation of intervention at scale - by ensuring e.g. that certain organizational conditions (identified perhaps in previous studies or reviews) are met - we can call it an experiment if we choose, but we now have to deal with a major and significant distinction between experiments that incorporate some sort of controls to test the counterfactual for the intervention, and other kinds of experiment.

Programme objectives and healthy ageing

Although several of programmes technology programmes included in this report as case studies are not explicitly concerned with healthy ageing, all the case studies have objectives which link with healthy ageing in one way or another. Several demonstrators aim to promote increased personal engagement with health among the general population or specific sub-groups within it. This is evident, for example, in *HealthSpace* and *dallas*. This is how one of published evaluation reports for *dallas* puts it: "Interactive person-centered digital tools and services offer a vehicle to promote a more citizen-led, self-care and a preventative health and well-being agenda"[j].

The same can be said of *HealthSpace*, which envisaged programme benefits (more empowered patients, improved health literacy etc) that align very closely with Wanless's idea of a population 'fully engaged' with its own health[k]. All the telehealth demonstrators take this same idea and apply it to particular sub-groups within the population, i.e. people with various LTCs. They are based on the conviction that the relevant technologies have the potential to empower individuals to improve their self-management (i.e. monitoring and responding appropriately) of their conditions. The same

[g] The evaluation used expenditure data from non-POPP areas as well as QoL data form the British Household Panel Survey to compare with measurements taken in the POPP areas.
[h] Originally integrated with *HealthSpace*.
[i] See below for more details.
[j] Devlin AM *et al* 2015
[k] This is most positive of the three health scenarios outlined in the 2002 Wanless Report.

underlying idea (individuals more actively engaged in maintaining their health) informs the 1st wave *Diabetes Prevention Programme* (*not* a technology demonstrator). The aim is to 'roll out' an individualised behaviour change intervention - a form of self-management to prevent the development of disease.

In other demonstrators, the link with healthy ageing is made by the way that service development or improvement is allied with a focus on preventive goals for older people who already have health problems or disabilities. The point of the innovation - whatever it is - is to intervene early with a view to preventing deterioration or loss of independence in at-risk individuals. So, for example, the *Combined Predictive Model* aimed to improve the targeting of preventive services for frailty. WSD saw digital technology as a means of enabling and stimulating the development of more integrated models of service delivery. POPP set out to extend the range of existing services by providing novel forms of early intervention. The various telecare programmes implement home monitoring as a way of maintaining independence - the ability to continue living at home rather than in an institution.

Last of all, and in a category of its own, there is the kind of holistic approach to public health interventions for lifestyle change that is found in the Healthy Towns Programme. The modality of intervention here was not service improvement for at-risk individuals or the promotion of individual engagement with health through individualised interventions but community and environmental change as an enabler for individual behaviour change.

Just as we can distinguish different modalities of intervention among the case studies, so can we also distinguish different target groups. *HealthSpace*, for example, was a minimally targeted programme - the general adult population. The various telehealth and telecare demonstrators target specific sub-groups within the population: people with LTCs and/or people with ADL-related needs. The Diabetes Prevention Programme targets individuals without diagnosed disease but at high risk of diabetes.

6. Recruitment and Participation

Recruitment

When researchers recruit participants or subjects into a trial with an experimental design, it is standard practice to include in the protocol targets for the number of people to be recruited into the study. These are people who consent to participate in a research study and know that this is what they are doing. The targets are set with a view to the requirements of the analysis of data collected as part of the study. The two policy experiments we have included with a technology focus had significant problems with recruitment, which is to say that they failed to meet their targets to an extent that compromised at least part of the analytical strategy built into the experimental design.

The target size for participants in the *Whole System Demonstrator* programme was fixed by RCT requirements i.e. the design was powered to answer key questions about the impact of remote care on health, service utilisation. Failure to recruit the target number of participants meant that the trial turned out to be underpowered according to its own design. More importantly, however, one of the inclusion criteria for participation was that individuals recruited into the trial should not already be using telehealth or telecare, i.e. they should be new adopters. These new adopters were to be recruited from randomly selected GP practices in the three implementation sites.

1. individuals who would benefit from using telehealth (and had one of the three diagnoses included in the protocol: COPD; heart failure; type2 diabetes)
2. individuals who would benefit from using telecare (because of difficulties with ADLs)
3. individuals who would benefit from using *both* telehealth and telecare

It was largely as a consequence of the requirement for new adopters that it proved impossible to recruit enough participants for the third of the three groups of study participants. As a result of the shortfall in recruitment for group 3, this particular intervention arm of the study was dropped, which meant that the study did not include subjects who had *both* health *and* social care needs. For obvious reasons, this adversely affected the ability of the demonstrator to explore the implementation of a more integrated approach to providing managing people with health and social care needs.

As with WSD, one of the main problems for *United4Health* was recruitment of participants. The total number of patients actually recruited was less than originally planned. The evaluators put this down to significant delays in the procurement of the necessary infrastructure (mainly telehealth equipment for use by patients with LTCs in the home), as well as unforeseen organizational changes in some of the deployment sites. The knock-on effect of the delay in procurement was to reduce the time available for recruitment: and the consequent shortfall in numbers led to the 'less than optimal' approach to the selection of participants[I].

Process evaluations may be affected by problems of recruitment as well as quantitatively-based outcome evaluations. In *HealthSpace* lack of interest amongst potential users of the technology had knock-on effects for the process evaluation. The evaluators could not find enough people who had

[I] Rasmussen, J. 2016.

spontaneously signed up to use the accounts. In the event their final sample was made up of people who consented to try the technology as part of a research project.

Participation, adoption, and uptake in the general public

Several of the other technology demonstrators used here as case studies had a different kind of participation problem. Their aim was not to recruit people into a research study, but rather to *scale up* a service, which in this context means encouraging or promoting the uptake of remote care technology by members of the general public in the relevant localities or implementation sites. The challenge was to expand the pool of users. The programmes had been given a definite time span - usually 3 years - during which they set out to grow the number of users by encouraging new adopters. The point to note here is that the number of new adopters is itself a measure of programme success. This may not be explicit in any of the available descriptions of the programmes, but it is clearly implicit in their goals.

When the evaluation of the *HealthSpace* programme ended in 2010, a total of 14 PCTs (out of 152) were participating. It was known that a similar web application was used by a large minority of Kaiser Permanante members in the USA. Could this kind of application (i.e. same functionality) be adapted for use in a different context - England - in such a way as to attract a reasonable number of patients? From the point of view of policy makers therefore, the main condition of success was uptake in the general population, and the evaluation made it clear that this condition had not been met. Between 2007 and October 2010, 172 950 people in the areas covered by the 14 PCTs opened a basic *HealthSpace* account. 2,442,215 people were invited to open an advanced account, and all of them had had (or were about to have) a summary care record created for them. Only 2913 people (0.13% of those invited) actually went through the steps required to open an account, which fell a very long way short of the 5-10% of the population anticipated in the original business case. In this case, it seems that not much serious thought had been given to the challenge of encouraging people registered as patients in the relevant GP practices to make use of a new web-based facility.

The same charge cannot fairly be laid at the door of the *dallas* programme, even though the results suggest that the difficulties of promoting the uptake of the new technologies among people in the general population who might benefit were seriously under-estimated. The challenges of 'scaling-up', of finding ways of creating new 'user communities', were central to the programme, which was set up with a relatively ambitious target for numbers of participants /users: it was hoped that the programme would reach 169,000 people (i.e. engage as new adopters) in the UK over 3 years[m].

Any final conclusion on the programme has to take into account the large and much-publicised shortfall in participation targets. The Scottish implementation site (Living It Up) had hoped to recruit 50,000 new users and managed 15,000. An interim evaluation report for the entire *dallas* programme reported 25,000 users at Jan 2015[n] (i.e. almost at the end of the programme. Although the evaluators do not suggest that this is actually decisive for their assessment of the programme, it is recognised as a major 'implementation challenge'.

[m] The target was ambitious for telehealth technologies, but not telecare technologies.
[n] Devlin AM *et al* 2015

One of the most important questions that the evaluations of both these programmes has to answer therefore is - why didn't they get closer to their recruitment target? Two kinds of explanation are offered by the evaluators. Firstly, there may be problems with the planning and implementation of the programme. So, for example, the *HealthSpace* evaluation is highly critical of the 'product' being offered - the web portal: *"Overall, patients perceived HealthSpace as neither useful nor easy to use and its functionality aligned poorly with their expectations and self-management practices..."*[o].

In the words of the evaluators, *HealthSpace* was hobbled by a 'flawed product'. Not enough thought had been given to how the technology was to work in the daily lives of its users. Although the quotation above refers to actual users, we can infer that the same problem occurred with potential users - why would they would want to use something if it appeared to be neither useful nor easy to use? It is hardly surprising therefore that one of their main recommendations for future projects was that the planning phase needed to give active consideration to "incentives, training and support for users."

It seems likely that one of the lessons that the planners of *dallas* took from *HealthSpace* was that they needed to involve potential users of new remote care technologies in the design of the products that would be made available. It turned out, unfortunately, that the time allowed for user-centred design processes was unrealistic, and this left too little time for roll-out, i.e. to increase the uptake of the new product. Hence the evaluators highlighted the problem of timescale in the implementation plan. Projects that incorporate both product development and national roll-out will require sufficient time for both[p].

The second kind of explanation for failure to meet targets for participation had to do with what the *dallas* evaluators call 'readiness'. In the *HealthSpace* evaluation this is described as a problem of 'flawed timing' - but it comes down to the same thing, at least in one respect. There is going to be little demand for the technology among potential consumers or users if they lack either understanding of the transformative potential of the technology (how it can make their lives better) or the relevant skills (how to make it work properly). In cases such as these, it takes time for the market to catch up with the technologies, especially if a large proportion of potential consumers belong to birth cohorts that less familiar with the technologies than younger cohorts.

This conclusion is further supported by the results of a Willingness-To-Pay study undertaken after the completion of the *dallas* programme, but by the same evaluation team and with a '*dallas*-like' cohort[q]. The paper begins by acknowledging that WTP research on consumer-facing digital health technology is still very much in its infancy. The headline result is that although consumers value mHealth solutions that "promote well-being, social connectivity and health care control, it is not universally embraced", which is to say that there was a high prevalence of zeros in both cohorts (and somewhat higher in the dallas-like cohort than in the general population). Moreover, and this is important, WTP was strongly associated with age (*older* people less willing to pay) and self-rated

[o] Greenhalgh T *et al* 2010
[p] Lennon MR *et al* 2017
[q] Somers C *et al* 2019

health (*healthier* people more willing to pay). To put it another way, there are problems with consumer readiness especially in those segments of the population that potentially could benefit most. The authors argue that *"the application of WTP in a digital health context demonstrates an economic argument for investing in upskilling the population to promote access and expedite uptake and utilization of such digital health and well-being apps".* Although the *dallas* evaluators see the problem of readiness as extending to other actors besides the general public (see below), the consumer is most emphatically included in their critique.

A third explanation for not meeting targets was offered by the HMHM evaluation[r] who stated that c *"The inherent level of complexity within HMHM, as well as the results of this evaluation, has shown that the benefits at a population level cannot be realised quickly*It is perhaps implicit in many of the evaluators' comments on the *HealthSpace* programme. Here though it is made explicit. The overall assessment of the HMHM programme is positive: *"there is good evidence of NHS resources being used more effectively and efficiently and hospital admissions being avoided through use of HMHM". But - as with HealthSpace and dallas - it is clear in the conclusions that achievement fell short of ambitions: "further scale up is required to reach population level gain".*

[r] Alexander H, 2018

7. Implementation

Complexity as a fundamental challenge

The evaluation reports for *HealthSpace*, the *WSD*, and *dallas* all highlight major implementation difficulties as factors that explain the lack of programme success. If there is a common underlying problem here, it is that the planners under-estimated the complexity of the systems within a newly introduced technology has to be made to work. The complexity moreover has a distinctive quality to it: the 'systems' in question incorporate people engaging in activities - which may be institutionalised (within service organizations) or not (in the case of service users or consumers) - and the nature of their engagement reflects their own understandings of what they are doing. They are 'socio-technical networks'[s]. We are in a domain that is proper to the social sciences.

The most egregious example is *HealthSpace*. The basic problem, as the evaluators saw it, was that the technology "was conceptualised more as a passive container for data than as [something] that could actively shape care". There was a failure to appreciate the implications of the technology for clinical practices (especially the interactions between clinicians and patients) and wider organizational routines. The implications had not been mapped out in advance and there were no formal plans for developing and embedding new modalities of interaction and new routines incorporated into the programme.

It is notable, however, that what looks like a very similar problem occurred in the *dallas* programme. This at least is what seems to be implied in the evaluator's comments on the pervasive problem of 'readiness'. What has to be emphasised is that the problem extended beyond the consumers or service users to embrace all aspects of the multiple and interlocking organizational networks which had to change in order for the use of the technology to be scaled up. *"Our results show that readiness issues have been ubiquitous across macro, meso, and micro levels and across sectoral boundaries: market, policy organization, professional and consumer"*[t].

In this case, the evaluators recommend that any future programme **needed to invest time and effort in assessing digital readiness of local health economies before embarking on innovation at scale**.

[s] A term used to highlight the relationship between new technologies and social practice. See Elzen B 1996.
[t] Lennon MR *et al* 2017

The NASSS Framework

NASSS is a heuristic framework which has been devised specifically to explain the successes, failures and changing fortunes of digital health programmes. In other words, it has an important retrospective application in programme evaluation. It also has a practical value, however, which is precisely the point argued by the HMHM evaluators. It can be used to:

- **to inform technology and service design at an early stage in technology development**
- **to identify innovations at a strategic planning stage that have little chance of achieving large-scale sustained adoption**
- **to address micro-level changes of individual adoption, meso-level challenges of organizational assimilation, and macro-level challenges of policy and regulatory environment**
- **to inform and support scale-up and roll-out of technology programmes**

NASSS is intended as an implementation framework as well as a framework for evaluation, and the advocacy on its behalf is explicitly critical of the approach taken by other implementation frameworks (including those used by the Department of Health in the UK). "Because e-Health technologies are typically introduced into a complex system in a turbulent and contested policy context", implementation frameworks designed around a "rigid and (apparently) systematic logic model" are likely to be counterproductive (Greenhalgh T et al 2017).

NASSS lays out a number of questions to be asked or challenges to be addressed across several domains. The main feature of the framework is that the questions (or challenges) are to be classified are simple, complicated or complex.

- **Simple phenomena/systems - straightforward, predictable, few components**
- **Complicated phenomena/systems - multiple interacting components or issues**
- **Complex phenomena/systems - dynamic, unpredictable, not easily disaggregated into constituent components**

The domains are: the condition or illness; the technology; the value proposition; the adopter system; the organizations; the wider context; embedding and adaptation over time.
The elaboration of the framework is accompanied by an analysis of case studies which suggests that implementers or planners tended to assume that the issues to be addressed were simple or complicated (hence knowable predictable and controllable) rather than complex. The conclusion is not merely that complexity in multiple domains poses a massive challenge to scale-up, spread and sustainability, but that

- **programmes which were complicated in multiple domains proved difficult, slow and expensive to implement;**
- **programmes which were complex in multiple domains did not become mainstreamed - or if mainstreamed - did not deliver key intended outputs.**

The evaluation of the HMHM programme in Scotland includes what is arguably a complementary recommendation, namely that *"planning for future developments should "**give consideration to how to reduce the level of complexity** of any domains that are currently rated as 'complicated' or 'complex'."*

They themselves made use of a framework for the analysis of complexity in rating different elements of the programme. This is Non-adoption, Abandonment, Scale-up, Spread and Sustainability (NASSS) Framework. They go on to recommend that *"further work should be undertaken to find ways of supporting an increase in the scale and pace of HMHM implementation, and that this work should make use of the NASSS framework".*

The point to note here is the view that a formal analysis of complexity needs to be built into the planning process for digital health demonstrator programmes (assuming here that the challenge of scaling-up is the main rationale for running the demonstrator in the first place).

Over-optimistic programme timeframes and inflexible timetables

The problem of timeframes - the misalignment between the time it actually takes to launch, implement and evaluate a programme and the time allowed for this in the implementation plan and timetable - is mentioned as a difficulty in virtually every case study included in this report. We have already seen it cited as contributory factor in explaining low numbers of participants in *dallas*. In this case the evaluators took the view that the programme should have been run for 5 years rather than 3. Too much was being compressed into 3 years: development of local implementation plans; negotiations with multiple partners; product development; roll-out etc. The *Summary Care Record* evaluation[u] highlighted the inflexibility of the timetabling built into the Department of Health's approach to project management; and recommended that **key milestones instead of being pre-defined should be locally negotiable[v].**

It is not only demonstrators with a technology focus that are affected by this problem, however. The *Healthy Towns* evaluation report concludes that the timeframe from the bid process to implementation was too short. They highlight the fact that implementation was rushed, which *"led to **limited consideration of the central administration processes that needed to be established to support local delivery structures**".*

As the quotation from the HMHM evaluation (see above) makes clear, it is fair to say that this recurrent problem of misalignment between the time it takes to implement a programme and the time allowed for completion is the obverse side of the problem of complexity. It is how the problem of complexity makes itself manifest.

[u] Included here because it was originally linked with the *HealthSpace evaluation*.
[v] Greenhalgh T, Russell J, 2010.

Nationally led, locally delivered?

One of the main dimensions of difference between the various case studies is the degree of freedom allowed to local implementation or deployment sites in deciding how best to achieve programme objectives that have been set at a national level. On the one hand, there are programmes which depend on a strong national template for local action and allow very little scope for local decisions about how best to tackle a given problem; and on the other, there are programmes which give local managers a great deal of discretion.

The *WSD* and the *Combined Predictive Model* fall unambiguously into the first category. The WSD required that the three implementation sites adhere closely to a pre-defined template for the design and delivery of the innovations they were testing precisely because it was being run as a controlled experiment. As we have seen, this caused major problems for the other programme objective - to use the technology as a way of facilitating or stimulating system redesign.

The *CPM* case is somewhat different. In 2006, the Department of Health commissioned work to develop two 'risk stratification tools' for the NHS in England (PARR and CPM). The aim was to help with the identification of people at high risk of unplanned hospital admissions. NHS Scotland and NHS Wales subsequently developed analogous models called SPARRA and PRISM. What is interesting about CPM is that it was supported – and promoted – by the Department of Health as a model that was developed nationally and could be applied locally without additional model adjustment to reflect local conditions. That is to say, the algorithm that CPM used was based on a nationally representative set of patient data and was intended to be applied locally 'straight off the shelf'. Subsequently, a number of PCTs decided to develop a 'local solution' to the problem by using local data for algorithm refinement. This could be then compared with the national tool for predictive accuracy[w]. Eventually, in 2011, the Department announced that it would not be commissioning a 'national upgrade' of either model (including the software used to run them). It looked as though the Government had come round to the view that local solutions might be better. The obvious conclusion was that the Government had decided "to promote 'plurality' in the information market"[x].

There are, on the other hand, several demonstrators that set out quite deliberately to leave a lot of room (as much room as possible?) for the development of local solutions to what was recognised a national problem. *HMHM* works with a group of 12 partners - a mix of NHS Boards and Health & Social Care Partnerships which competed to secure funding for participation in the programme. Partners could determine locally what population groups would be offered HMHM, and also - locally - identify through user engagement which technologies would be adopted. This is not the same as the user-centred product design process used in *Dallas*. The partners were to deploy existing technologies, but they would consult potential users before making procurement decisions. The problem was specified at a national level - how to scale up the use of remote care technologies among the general public. It was left to local partnerships to decide on the target groups/care pathways as well as the technologies (and of course the two decisions go hand in hand). It is not really possible, however, to say how much this localist approach contributed to the qualified success

[w] See eg, https://www.wirralintelligenceservice.org/media/1056/predictive-model-_emergency-admissions_-v3.pdf
[x] https://www.nuffieldtrust.org.uk/news-item/parr-is-dead-long-live-predictive-modelling

of the programme. The most we can say is that it takes steps to avoid problems that occur when no account is taken of the need to adapt project proposals to local circumstances.

The localist approach has been taken a stage further by the Social Care Digital Innovation Programme. Whereas HMHM ask local groups to determine which population groups would be encouraged to adopt a particular range of technologies (i.e. technologies for home and health care monitoring), the SCDIP appears to give local manager one more degree of freedom. Funding for the programme has been broken down into 2 phases: a discovery phase and an implementation phase. The discovery phase was introduced on the suggestion of the evaluators to "facilitate the exploration of a range of factors that contribute to more successful implementation, scalability and sustainability by identifying and responding to an evidenced need". Successful applicants received funding to "determine the project needs, test the research challenge, explore solutions and ensure user needs are met by project proposals"[y].

Local freedom without adequate central guidance can itself be problematic, however, as the *Healthy Towns Programme* shows. Like POPP, this programme gave local partnerships freedom to develop their own service responses to a challenge that was specified out nationally. The problem that arose with the Healthy Towns Programme was the level of detail provided nationally regarding the type of service response/local programme that was expected. The government wanted examples of 'whole system' approaches to tackling obesity, but did not spell out in sufficient detail what was required of local programme for it to qualify as an example of a 'whole system' approach[z]. For the evaluators the inadequacy of central guidance on this matter was the main reason why only one out of the 9 funded local programmes actually delivered an example of a 'whole system' approach.

Locally led, locally delivered

The obvious alternative to nationally led and locally delivered is locally led as well as locally delivered. The trouble here is that it is not always possible to be sure that the programme in question would count as a demonstrator. If scaling-up becomes incorporated into strategic planning as 'business as usual' and evaluation is transformed into (i) monitoring to check on the delivery of the strategy (ii) including regular assessment of service quality (asking users what they think), then it ceases to bear most of the hallmarks of a demonstrator programme - even though the relevant strategic objective is to promote the uptake of this or that new technology among the general public. What has happened in Liverpool since the completion of the *dallas* demonstrator is a case in point. Scaling-up is presented as part of a longer term strategy that includes medium-term (5-10 years) as well as long-term goals (10+ years)[aa]. At the end of 2017 the CCG announced to expand its efforts to scale up the local telehealth services. We should note here that much of this was seeded by *dallas*, which serves to exemplify the discussion in the following section about the timeframes for evaluation.

[y] See appendix for more details.
[z] Cummins S *et al* 2016
[aa] Healthy Liverpool the Blueprint, Liverpool CCG 2015

Healthcall in the NE of England is similar[bb]. "Health Call enables professionals in the region to implement digital care at scale. Health Call supports NHS organisations to design, develop and implement digital health services at speed, with content also made available to the other organisations in the UK". In this case, 6 CCGs have come together to develop relevant work that no one of them has the capacity to take forward on its own. As with Liverpool, this is more like a plan of work than a technology demonstrator. They want to develop their own 'solutions' and set their own timescales, and if a pilot is called for - i.e. to test an innovation - they can work out a collaboration with the regional Academic Health Science Network.

[bb] https://www.nhshealthcall.co.uk/

8. Evaluation

Evaluation Guides

There are some robust sets of guidelines or toolkits for evaluating technology programmes in healthcare.

1.The e-Health Implementation Toolkit (e-HIT) was used in the evaluation of the dallas programme considering 4 different aspects of the implementation of e-Health:

- coherence - how do individuals make sense of new ways of working?
- cognitive participation – what was the participant engagement and 'buy-in'?
- collective action - how was the programme put into practice?
- reflexive monitoring - what procedures were put into place for monitoring?

2. MAST (Model of Assessment for Telemedicine) was used in the United4Health programme. It builds on and adapts the Core Model of the EU Network for Health Technology Assessment to the specific purpose of telemedicine or telehealth. (See United4Health)

3. Non-adoption, Abandonment, Scale-up, Spread and Sustainability (NASSS) Framework is a heuristic framework which sets to explain successes, failures and changing fortunes of tech-supported programmes. In addition it may be used to frame throughout the project

- At an early stage in technology development to inform tech and service design
- In strategic planning to identify innovations that have little chance of achieving large-scale sustained adoption
- In technology implementation projects to address micro-level changes of individual adoption, meso-level challenges of organizational assimilation, and macro-level challenges of policy and regulatory environment
- To inform and support scale-up and roll-out of technology programmes.

NASSS lays out a number of questions to be asked or challenges to be addressed across several domains. The main feature of the framework is that the questions (or challenges) are to be classified are simple, complicated or complex.

- Simple phenomena/systems - straightforward, predictable, few components
- Complicated phenomena/systems - multiple interacting components or issues
- Complex phenomena/systems - dynamic, unpredictable, not easily disaggregated into constituent components

The timeframe for evaluation

Just as timeframes can be a problem for implementation, so can they be a problem for evaluation. Is sufficient time being allowed for evaluation? The answer will depend on the criteria for success and the relevant measures of impact. The local impact evaluations conducted under the broad umbrella of the Healthy Towns Programme were looking for changes in behaviour that could be attributed to the interventions implemented in each area. What is a reasonable amount of time post-implementation (itself a very time-consuming business) to allow for effects such as these to become apparent in locally collected data? It seems clear the 3 years from launch to programme completion was not enough. In *dallas*, there was a big emphasis on lifestyle change in the two locality-based programmes, LIU and Mi. It was hoped that the new technologies – or some of them anyway - would enable and promote lifestyle change in the 50+ population. Was enough time allowed post-implementation to register *sustained* change?

The same issue, albeit in a different guise, emerges in demonstrators that have the scaling-up of an innovation as a primary programme objective. In programmes such as *dallas* or *HMHM*, 'hard' targets were for numbers of new adopters or users of a new service. How large a shortfall from the chosen target should be read as a sign of failure? In what circumstances should the argument that more time was needed - that the cut-off date was arbitrary - be accepted?

The report for *HMHM* goes some way towards resolving this issue by the way it distinguishes between (i) outputs and outcomes (ii) and short-term (yr 2) , medium-term (yr 3) and long-term outcomes (2030). Outputs include items like no's adhering to treatment or coaching programmes. Outcomes include higher % increase in condition control and more timely appointments. Outputs include items like no's adhering to treatment or coaching programmes. Outcomes include higher % increase in condition control and more timely appointments. They also make sure that they can provide clear-cut answers to questions about spread as well as scale-up. One of the corollaries of this is that the 3-year evaluation report is not, strictly speaking, a *final* report on a *completed* programme. Initial funding for local programmes comes to an end after 3 years, but this does not mean that the larger umbrella programme also comes to an end after 3 years. Three years is simply not long enough to assess the kinds of change that are being sought (e.g. spread as well as scale-up), and it may be not that easy to say how long is long enough.

Something similar can be seen in the *United4Health* project. In this case, the fact that the project was an experiment with fixed funding over a 3 year period meant that some of the evidence claimed in the evaluation report as signs of success (that is it feasible to deploy telehealth services at scale) emerged outside the agreed timeframe. What was claimed here as evidence of success is the fact that telehealth services as implemented in the project continued operating after project completion in nearly all the deployment sites and that plans had been drawn up in these same sites to continue with the scaling-up services. The scaling-up of telehealth service is an ongoing process that does not fit very easily within the relatively short timeframes required by this kind of programme funding.

The same point underlies of the recommendations of the WSDAN evaluation report. Strategies for scaling-up have to find ways of combining continuous monitoring and quality improvement processes with a governance structure that ensures that definite programme goals are achieved[cc].

Controlled experiments

In 2010, the same year as the publications of the results of the *HealthSpace* evaluation, the lead author of the report, Trish Greenhalgh, wrote a paper with the provocative title - *Why do evaluations of eHealth programs fail?* The paper begins by pointing to a specific failure in the evaluation of the associated *Summary Care Record* programme. The attempt to set up something like a controlled experiment - with a step-wedge design - was eventually abandoned in favour of a mainly qualitative case study[dd]. Her diagnosis turned on the problems of evaluating programmes implemented in situations where goals and criteria of success are 'contested' by different stakeholders. We do not have to agree with all the details of her analysis to concede the point that experimental designs may not be appropriate in certain kinds of circumstances, especially if successful implementation requires changes in practice in a complex system - in this case, a 'socio-technical network'.

Her analysis was to some extent borne out by the difficulties that arose for the WSD evaluation where local managers and the Department of Health had different views about priorities (see Box 1). The requirements and constraints imposed by running a controlled experiment may clash with the need for ongoing adaptation and negotiation between different stakeholders that is likely to characterise a pilot programme which aims to implement a new model of service delivery. The point being made here, however, is that the *programme* failed to achieve one of its main objectives. Would it be fair also to say that the actual *evaluation* failed? If we think that the evaluation failed to produce robust evidence for the positive impact of telehealth and telecare *because* the attempt to run an RCT in this situation was ill-conceived, then presumably it is fair to say that the evaluation failed. Moreover, we might wonder whether the evaluation really provided useful answers to one of the questions it was meant to ask - how might the challenges that providers encounter in implementing and scaling-up telehealth and telecare services be overcome? The fact that such challenges were not altogether overcome in the *dallas* programme counts to some degree against the usefulness of the WSD evaluation. It was unable to come up with a recipe for success - though it remains an open question (not answered by the evaluators of either programme) whether this is reasonable expectation. This is not to say that *dallas* did not build on lessons from previous failures: it heeded the warning from the WSD about attempting to run an RCT, and responded to the 'flawed product' criticism made of *HealthSpace* by incorporating user-centred design processes.

The attempt to run a controlled experiment in the *United4Health* project also ran into problems, but in this case they were rather different. The evaluation report cites various issues that arose as a result of running this kind of experiment in several countries, namely the extent to which local conditions (i.e. the resources available to services at different sites as well as key aspects of their operational procedures including data collection) varied from country to country. Indeed the report

[cc] Greenhalgh T, Russell J, 2010

[dd] Something similar happened with the Healthy Communities Challenge Fund evaluation. The original plan to generate robust evidence of impact was modified in favour of a mainly process-type evaluation of project development and implementation.

argues that the failure to demonstrate positive outcomes for the cost-effectiveness of telehealth deployment was due mainly to significant variations across the different sites in factors likely to affect the outcome such as: number of patients that could be supported; costs of telehealth solutions; existing care pathways; and general support infrastructure. To put it another way, the project team was unable to ensure that the conditions were sufficiently homogeneous across the different deployment sites for the trial to work as a controlled experiment. With the same number of participants across multiple sites in a single country, they might well have received a more positive result.

It would be a mistake, however, to suppose that the use of an experimental design has failed in *all* the technology case studies included in this report. They were used in local iterations of the *Combined Predictive Model* risk stratification tool, and current guidance on this matter recommends various approaches to the incorporation of controls[ee]. Some of the *NHS Test Beds* are being run as controlled experiments, in this case, moreover, with full awareness of the kinds of issue that have bedevilled past evaluations of service innovations involving digital health technologies.

Heterogeneity in interventions as a challenge for evaluation

The *Partnerships for Older People Projects* (POPP) demonstrator programme and the *Healthy Towns Programme* (aka Healthy Communities Challenge Fund or HCCF) shared one very important feature. They were set up in a way that more or less guaranteed a large and heterogeneous set of interventions for the evaluators to take into account. The POPP programme involved 146 core local projects (many different modalities of service delivery) in 29 different sites. Individual projects were developed and selected according to local priorities subject only to the overall aim of the programme - that they should promote the well-being of older people and prevent avoidable hospital admissions among the same population. Despite this heterogeneity, there was a national evaluation which managed to assess impact on (i) older people, (ii) joint working in the locality and (iii) expenditure. What enabled the evaluation to tackle the heterogeneity was a relatively simple typology of types of project and a small number of outcome measures that could be applied across the board or to specific types of project. The evaluation worked in the sense that it could provide answers to most of the key questions posed as part of programme. It could even differentiate between types of project, i.e. which worked better[ff].

The Healthy Towns Programme was on the face of it very similar. It ran in 9 separate urban areas, selected to be representative of different types of urban areas found in the country. The localities were free to develop a locally-specific programme of interventions subject only to the overall aim of the programme - to test 'whole town' approaches to environmental barriers to reducing obesity in the general population (by increasing physical activity and improving. As with POPP, one of goals of the Healthy Towns Programme was positively to encourage a wide range of different types of intervention, and in this they succeeded - with 306 separate projects. Although impact evaluations were conducted locally in all the sites, the guidance provided by the Department of Health on the commissioning and conduct of these evaluations was thought to be inadequate by many local stakeholders, and in the event no attempt was made to unify the local impact evaluations into an

[ee] *Next steps for risk stratification in the NHS*. NHS England 2015. Summarising evidence to date the report concludes *inter alia* that available tools may actually increase costs rather than reduce them.
[ff] Windle 2009

overall assessment of programme impact at national level. The national evaluation was steered away from impact to process. *"The sheer heterogeneity of Healthy Town interventions makes it difficult to synthesise findings across the whole programme, reducing the opportunity to evaluate programme level impact"*[gg].

Heterogeneity is, one might say, a predictable cost of allowing local freedom in determining the content of projects. How much heterogeneity is too much is hard to say, but this hardly removes the requirement to make a considered judgement on this point.

Accounting for complexity in evaluation

Where complexity is a feature of an innovation or intervention, it poses problems for evaluation, and these have been much discussed in the research literature. MRC guidance on the evaluation of complex interventions includes advice on both impact evaluation (appropriate experimental designs for determining practical effectiveness) and process evaluation. Process evaluations are here seen as *complementary* to an impact evaluation, and not as a substitute (used to assess fidelity and quality of implementation, clarify causal mechanisms etc)[hh]. This prompts the question - what are evaluations of non-experimental demonstrators trying to achieve?

For the technology demonstrators, an answer to this question may perhaps be found in the theory that underlies some of the heuristic tools intended to help with the analysis of the process of developing and implementing an innovation.

Normalization Process Theory (NPT) underpinned the e-Health Implementation Toolkit (e-HIT) that was used in the evaluation of the *dallas* programme. NPT started life as an applied theoretical model to assist clinicians and researchers to understand and evaluate the factors that inhibit and promote the routine incorporation of complex healthcare interventions in practice. It was subsequently generalised to a theory of normalization processes that proposes a working model of implementation, embedding and integration in conditions marked by complexity and emergence. *"The theory focuses on the work of embedding and of sustaining practices within interaction chains, and helps in understanding why some processes seem to lead to a practice becoming normalized while others do not"*[ii]. NPT underpinned the e-Health Implementation Toolkit (e-HIT) was used in the evaluation of the *dallas* programme. The idea is to consider 4 different aspects of the work involved in the implementation of e-Health:

- coherence - how do individuals make sense of new ways of working
- cognitive participation - participant engagement and 'buy-in'
- collective action - how was the programme put into practice?
- reflexive monitoring - what procedures were put into place for monitoring.

The NASSS framework we have already mentioned. As well as being an aid to designing and implementation it has an analogous role to play in evaluation. As with NPT the retrospective use of NASSS framework is intended to help pinpoint reasons for success or failure in scaling-up. We have

[gg] Goodwin DM *et al* 2013 p111
[hh] MRC 2006; Moore G *et al* 2015.
[ii] May CR *et al* 2009

also to bear in mind that one of the main messages of NASSS framework is that too much complexity is a recipe for failure. Guidance on the evaluation of complex interventions becomes irrelevant if the interventions are 'too complex'.

One of the most interesting features of the *HMHM* programme is the approach it takes to evaluation. As well as NASSS (which is here used primarily to give an overall rating of complexity), it uses another theory-based heuristic, namely Contribution Analysis (CA). The point of CA is to help with problems of causal attribution in complex real-work stings where it is rare for a single action to be solely responsible for bringing about a given effect. It works "by gathering evidence to support an agreed theory of change (or logic model) that can be refined over time, thus generating credible claims that link a range of activities to observed results"[jj]. In other words, it is a way of getting at causal mechanisms when - for whatever reason - the kind of approach recommended by the MRC (experimental design + process evaluation) does not seem feasible.

Diversity in unity: the NHS Test Beds

If HMHM is interesting for its approach to evaluation, then the same is to be said of the Test Bed programme, which aims to "provide a *novel* (my italics) approach to test innovations in health systems, by tackling three well established barriers to the uptake of innovation"[kk]:

- innovations are often implemented in isolation from each other and the infrastructure in which they function
- there is a comparative lack of robust evidence about their real-world effects, as opposed to experimental or research settings
- innovations are often introduced on top of existing working practices and infrastructure, rather than in conjunction.

The unifying question for the Test Beds is - have NHS-innovator partnerships worked and why and if not, why not? Each Test Bed, however, takes a different approach to answering these questions. Hence the techniques being used are varied and include a randomised control trial, matched control studies, observational studies, feasibility studies and soft system methods. The end product - the output - should provide a considered "view of the value of each innovation to the NHS. This will be accompanied by an understanding of the conditions that have produced particular outcomes and why, and a better sense of what type of evaluation approaches best suit different types of interventions". Although there are as yet no published evaluations, the programme has published a handbook on evaluation which explains the choices made about evaluation methodology in the 1st wave projects.

What is notable in the Test Beds is that the different elements in the programme do not stand or fall together. Failure, however that is to be determined, in one of the Test Beds should have no bearing on the success of the others. This is in marked contrast to programmes in which the different elements (or sites) are conceived of as a piece and evaluated as a single unified programme providing evidence for the answer to *one* question or contributing to *one* objective.

[jj] Alexander H 2018.
[kk] Galea A *et al* 2017

9. Questions and Lessons

We have noted and commented on many different kinds of demonstrator and in doing so have tried to highlight major dimensions of difference. Most obviously there is the difference between demonstrators with and without a specific technology focus. Most of the comment - and the theory - we have included arises out of the many problems that have beset various technology demonstrators in this country and elsewhere. Whether or not technology demonstrators pose special problems or add an extra layer of complexity to an already complex situation it is hard to say. One very general conclusion would be that complexity and ambition increase the risk of failure.

Questions

In considering the design, implementation and evaluation of a demonstrator project for healthy ageing the following frame is useful.

1. **What is special about technology innovations and demonstrator programmes with a strong technology or service focus?**

2. **How do we balance complexity, experiment, and purpose?**

3. **What are the types of learning**
   - **Learning from failure vs learning from success**
   - **Learning specific to remote care/socio-technical networks**
   - **Formative vs summative evaluations**

4. **How do we design, implement and evaluate a demonstrator project, in particular its impact on ability to deliver innovative services**
   - **Examples of sets of guidelines or toolkits for evaluating technology programmes in healthcare (e-HIT and MAST)**
   - **Lessons from Normalization Process Theory (NPT)**
   - **Non-adoption, Abandonment, Scale-up, Spread and Sustainability (NASSS) Framework**

5. **What is the feasible scale of a proposed demonstrator to be delivered within 3 years?**

6. **What mechanisms are required to foster learning and innovation between sites?**

7. **How do we approach technology readiness, and the implications for moving a technology to scale?**

1. Policy experiment or demonstrator?

It is important to ask whether it is necessary or desirable to run an experiment, and if not, to be clear about the purpose of the demonstrator and the nature of the evaluation, including the appropriate criteria of programme success and failure. What is expected from the evaluation in the case of non-experimental demonstrators?

2. Off-the-peg technologies?

No firm conclusion can be drawn about including the development or design of new technologies as an integral part of a technology demonstrator programme. This was a cause of difficulty in *dallas* - because of the impact on timeframes - but the evaluators did not regard it as an insuperable problem. Indeed they wanted to see more effort put into finding optimal approaches to executing user-centred design processes.

3. What degree of local control is desirable?

The case studies we have selected illustrate different approaches to the balance between national objectives and local implementation. The recent history of technology demonstrators suggests a shift towards localism of one kind or another. Adaptation to local conditions is increasingly seen as a precondition of success.

4. Ambition v. realistic scale?

Both WSD and *dallas* were very ambitious programmes. In retrospect it seems that they were trying to do too much too quickly. The problem was not the ambitions as such - these point to challenges which are still real and significant , e.g. large-scale care system redesign, the transformation of local health economies etc - but the expectation that they could be solved all in 'one go' by a single time-limited programme.

5. The question of heterogeneity in interventions?

Control over the quality of an intervention (fidelity to an agreed model) and the conditions of delivery are important for an experiment. The decision to loosen that control and allow for multiple interventions with different target populations can create a serious headache for evaluation.

6. Timeframes?

Perhaps the main lesson that is reiterated again and again in the examples we have chosen is that scaling-up is very a complex and time-consuming business. Because planners repeatedly under-estimate the complexity of what they are undertaking, implementation very often (in our examples - always) takes longer than expected. Challenges may be exacerbated by inflexible timetables. Evaluations are also constrained by the time periods they cover. Choice of outcomes is obviously important here, and if scaling-up is a primary objective, then most evaluations agree in thinking that 3 years is too short for an evaluation period.

7. Demonstrators vs core business?

There is a fine line between running a demonstrator and incorporating scaling-up as an objective within a strategic plan. Many CCGs have now developed organizational structures and plans that reflect their commitment to realising the potential of certain kinds of new technology. This commitment, together with continuing innovations and improvements in health technology and increasing public familiarity with digital technology, is transforming scaling-up into routine organizational business. They still have to find ways round implementation challenges and make decisions about what to test and how, but in some cases at least this has become part of their core business.

8. Learning from failure and incremental learning?

It is very obvious that successive technology demonstrator programmes have learnt from the failings of predecessors. To take the obvious example, *dallas* was launched in 2012, after the WSD failed to establish a 'business case' for telehealth, and was at pains to avoid some of the pitfalls and problems that evaluators had highlighted in previous programmes. What *dallas* did not do was come up with a definitive recipe for success in overcoming implementation challenges. It can be regarded therefore as another staging-post in a long, drawn-out process of incremental learning. If this point is accepted, we need perhaps to ask (a) whether or not it is reasonable to look a definitive recipe for success (b) how important it is to use demonstrators as ways of trying out well-specified recipes for success.

Lessons

A programme that focuses on the potential of new technologies to contribute to the promotion of healthy ageing, requires it to be explicit about the modalities by which change is to be effected, to emphasise the importance of novel framing of challenge and to place people at the centre not technology.

Some broad lessons identified from the review :

- Produce a clear vision
- Identify measures of success
- Incorporate specific targets
- Recognise complexity
- Consider "technology readiness
- Recognise trade-offs
- Facilitate integration
- Support leadership
- Talent pipe-line
- Build in adaptation

1. Vision: clarity about the purposes being served by running a demonstrator entails clarity about the kind of knowledge or know-how that is expected from the evaluation.

Ensure all key players understand and share a sense of the vision. The challenge is how to get an intervention or innovation 'up and running' or embedded in routine practice in a way which can be expected to produce the desired outcome at a population level. The problem is often presented in terms of 'scaling-up' - of "moving from a local project to one that is 'business as usual'", or from ideal conditions to real-world conditions. Some of the programmes were designed and run as experiments - they set out to determine whether or not an innovation or intervention delivered the expected benefits, and they used an appropriate methodology.

The rationale for piloting an initiative without running it as an experiment is that scaling-up is recognised as challenging. Successful implementation at scale cannot be taken for granted. Nor can the means of achieving it. What both kinds of demonstration programme have in common is that they cannot fulfil their function without some form of evaluation.

The basic decision to consider in planning a demonstration programme is whether it is necessary or desirable to run an experiment, and if not, to be clear about the purpose of the demonstrator and the nature of the evaluation, including the appropriate criteria of programme success and failure. This is especially true now at a time when many CCGs are incorporating the challenges of scaling-up the use of new technologies into the routines of strategic planning business and incremental organizational learning (i.e. core business rather than demonstration programmes).

2. Identify measures of success

Demonstration programmes can fail in different ways because they have different purposes in view. Experimental demonstrators can fail to produce robust evidence of positive impact or it may prove impossible to implement the original experimental design.

- The *Whole* System *Demonstrator* and *United4Health* ran into both these problems. A programme that has implementation at scale as its primary aim fails if the innovation is not up and running by the end of evaluation period or if participation targets have not been met.
- *HealthSpace* and *Dallas* failed to the extent that the levels of public participation they actually achieved fell way short of their ambitions or expectations. They failed in terms of their primary objective.
- *Healthy Towns Programme*, which set out to test 'whole systems' approaches to obesity reduction, and in the event succeeded in implementing a whole system approach in only one of its 9 demonstration sites also fell short.

3. Incorporate specific targets: for longer-term goals like spread, sustainability and scale-up

Measures of success need to be integrated within the project design:

- Scale-up – within the setting
- Spread – transfer to new settings
- Sustainability – being maintained long term, adapting as required

Reasonable scale-up for a given condition will drive forward population health benefits at scale but 3 years is too short.

4. Recognise Complexity: avoid assumption that the issues to be addressed are either simple or *complicated* (hence knowable predictable and controllable) rather than *complex*.

Assess the nature and extent of complexity in the programme and ensure that emergent and adaptive measures are used to address these issues

- *complexity* in multiple domains poses a massive challenge to scale-up, spread and sustainability.
- programmes which were *complicated* in multiple domains proved difficult, slow and expensive to implement.
- programmes which were *complex* in multiple domains did not become mainstreamed - or if mainstreamed - did not deliver key intended outputs.

A recurrent theme for all the technology demonstrators is that planners seriously under-estimated the complexity of the systems - the socio-technical networks - within which a newly introduced technology has to be made to work. Failing to account for complexity is one of the main reasons why attempts to implement at scale so often fail to keep to programme timetables or achieve participation targets.

The recommendation of the HMHM programme evaluation, that a formal analysis of complexity needs to be built into planning process for future digital health programmes, deserves serious consideration.

5. Consider "technology readiness": define the circumstances in which it would be useful to conduct an assessment of 'technology readiness' as part of the planning process

Whether or not technology demonstrators pose special problems or add an extra layer of complexity to an already complex situation it is hard to say. Lack of 'technology readiness' in the socio-technical network being targeted for implementation has been proposed as a possible explanation for programme failure.

6. Trade-off: acknowledge there is a trade-off between programme ambition and scale, on the one hand, and a manageable timetable, on the other.

The challenge of complexity is a challenge for timescales. Evaluations are constrained by the time periods they cover, and it is possible for programmes to be judged prematurely as failures - because the initial timeframe for achieving particular targets was unrealistic. Inflexible timetables may not allow time to respond to emergent problems. The reports we saw agreed in thinking that **3 years is not enough time for the evaluation of a large and ambitious programme. Our recommendation is a 5 year plan.**

7. Integration: findings from separate evaluations of distinct interventions need to be integrated so as to provide an easement of overall programme impact

Control over the quality of an intervention, fidelity to an agreed model, and the conditions of delivery are important for an experiment. The decision to loosen that control and allow for multiple interventions with different target populations can create difficulties for evaluation.

8. Leadership: directives for both ownership and local leadership built into initial plan

The question of how local ownership may be built into a national programme with overarching objectives requiring explicit consideration. In addition, it is necessary to establish overall leadership since complex programmes often suffer from outsourcing of control and coordination. The case studies illustrate different approaches to the balance between national objectives and local implementation. The recent history of technology demonstrators suggests a shift towards different forms of local 'ownership' for the implementation of national programmes.

9. Talent pipe-line

A talent pipe-line is essential to retain knowledge and expertise in the workforce developed over the course of the programme.

10 Adapt: build in adaptation as the programme evolves.

Collect and reflect on emerging data, be creative, but control over ambitious growth, since projects that evolve organically are vulnerable to over-ambitious extension and scope creep.

Part 3 Evaluated Programmes

10. National demonstration programmes with a strong technology focus

1.HealthSpace

Type: early implementation pilot /demonstrator for web-based health application

Programme duration: From 2007 until it was eventually axed in 2013. Although it was originally integrated with the *Summary Care Record* pilot, the two programmes were uncoupled in 2010 (locally and nationally). An upgrade to the functionality of HealthSpace had been planned for 2010 (including a facility to order repeat prescriptions online), but was not approved by the Department of Health as it was seen as too risky. The independent evaluation covered the period from the commencement of the programme till 2010.

Estimated cost: £8 million for 2007-10.

Description: an internet-accessible personal organiser onto which users of health services could enter their own personal health data (such as blood pressure) and plan health appointments. People who took up the option of registering for an *advanced* HealthSpace account could also gain secure access to their Summary Care Record (SCR) and email their GP using a function called Communicator. The advanced account (web application) provided users with personal information about their health (including laboratory results, medications, past clinic visits and key diagnoses) and enabled them to manage their interactions with health services. The advanced account required cooperation with the SCR project.

Evaluation methods: basic data on new accounts and qualitative assessment of user problems and issues.

Documentation: Report and peer-reviewed journal papers available.

2. The Whole System Demonstrator (WSD)

Type: experiment/trailblazer for telehealth and telecare

Programme duration: From 2008 to 2011. This coincided with the evaluation period.

Estimated cost: £30+ million

Description: The core component of the programme was a Randomised Controlled Trial designed to assess the effectiveness - and cost-effectiveness - of remote care technologies with a view to producing the kind of evidence that would persuade clinicians to incorporate them routinely in their practice. The aim was to provide 'credible evidence that comprehensive integrated care approaches combined with the use of advanced assistive technologies, both benefit individuals and deliver gains in cost effectiveness of care'.

- Telehealth technologies - based on the remote exchange of data between a patient and a health care professional - would be supplied to individuals living in the community to assist in the management of an existing long-term condition (LTC).
- Telecare technologies would be supplied to individuals living in the community with a view to providing remote automatic monitoring of an individual's personal health and safety in the home environment.

The WSD also had a second objective, which was to act as a demonstrator/learning pilot for the use of advanced assistive technologies in driving 'whole systems redesign'. In this context, *whole systems redesign* refers to an array of changes in organisation and practice that aimed to produce more integrated – more joined-up – and hence more effective health and social care services. It was conceived, in other words, as a response to perceived failings in the health and social care 'system'. There would be improved sharing of information and better co-ordinated care plans; and, the traditional model for delivery of care would be supplanted by encouraging care in the home through increased use of Community Matron case management and self-care management. Telehealth for people with LTCs and telecare for people with social care needs were introduced in the WSD implementation sites to facilitate these aims.

We can think of this second part of the programme as asking a number of subsidiary questions:

- What challenges do providers encounter in implementing and scaling-up telehealth and telecare services? And how might they be overcome?
- Can remote care be used to stimulate whole system change? And more broadly, can it be used to change the current forms of service delivery for patients?
- What organisational factors facilitate or impede the sustainable adoption and integration of telehealth/ telecare?

Target groups: The selected population groups corresponded to the functions of the technologies to be provided and tested: individuals with LTCs - in this case either COPD or diabetes or heart failure - for telehealth; individuals with fairly serious problems with ADLs for telecare. 6000+ individuals were recruited into the trial. The Demonstrator was run in 3 implementation sites.

Evaluation methods: RCT + process evaluation. Unlike most other evaluations of similar technologies used in the NHS up to that time, the trial was not designed primarily to assess the effectiveness of a specific type of TC or TH technology. The analysis was conducted at the level of the technology type (i.e. TH or TC), and no plans were included to conduct analyses of individual technologies.

Documentation: Published report and peer-reviewed journal papers available.

3. WSD Action Network

Type: multiple learning pilots for telehealth and telecare

The WSD Action Network was made up of 12 sites that were not included in the Whole System Demonstrator project (3 sites only) but had all been applicants to the WSD and were interested in developing their own remote care programmes. The development activities of the network members were subject to an independent evaluation covering the same time period as the WSD itself (2008-11). Although these activities did not form part of a single centrally-led programme (12 separate local initiatives), they were part of a single centrally commissioned evaluation.

Evaluation methods: Unlike the WSD, this was not an 'impact assessment'. Nor was there any attempt to compare achievements with objectives. The focus was quite explicitly on process, i.e. progress made in adopting telehealth and telecare and learning about implementation challenges.

Documentation: Published report available.

4. dallas (Delivering Assisted Living Lifestyles at Scale)

Type: demonstrator/learning pilot for telehealth and telecare

Programme duration: Dallas was launched in 2012, after the WSD failed to establish a 'business case' for telehealth. It ran as a demonstrator pilot till 2015, with the evaluation covering the same period.

Estimated cost: £37 million

Description: The main aim of the programme, which was presented as the culmination of work undertaken by the Assisted Living Innovation Platform (ALIP), was to promote the roll out of digital health and wellness products, services and systems in order to:

- 'show how assisted living technologies and services can be used to promote wellbeing, and provide top quality health and care, enabling people to live independently and improve quality of life' ;
- 'unlock new markets in social innovation, service innovation and wellness, enabled by technology, and show that technologies and services can be made available at a sufficient scale and cost to enable independent living.'

Dallas was a very complex programme, and was set up with four distinct organizational components:

- *Living it Up* (LIU) and *More Independent* (MI) were locality-based projects to implement new technology-enabled services in two parts of the country: Liverpool (MI) and Scotland (LIU).
- i-focus was "an overarching dallas community, consisted of three workstreams that aimed to bring technical and strategic coherence, cross-sector collaboration, consumer-focused innovation and real commercial scale to the programme". It established the Digital Health and Care Alliance (DHACA) as a national membership-based digital health interoperability community.
- Year Zero - industry-led consortium focused on product development and information governance. One of the original ambitions for Year Zero was to deliver a commercial 'e-Redbook' - an electronic Personal Child Health Record.

Unlike the WSD, which used pre-existing technologies, the intention was that Dallas would use a variety of approaches to user-centred design. It was also at pains to differentiate itself from earlier demonstrators (not WSD) that focused on single digital tools or systems (i.e. one product at a time). The aim was "to deploy a broad portfolio of digital tools and services" and it was building on these earlier single-product initiatives to become "the next stage toward deployment of such technologies for health and well-being at scale in the UK".

Target groups: LIU took a very inclusive view of the target populations for the programme: (i) general population (ii) people between 50 to 70 years of age who were active and healthy (iii) people between 50 to 75 years with or at risk of Long Term Condition (iv) Over 75 years with LTC or Frailties.

Evaluation methods: dallas was not designed as a trial or experiment. The whole programme was quite explicitly conceived as a demonstrator/learning pilot rather than as an investigation of the effectiveness of digital technologies in health care. For this reason there was a different approach to evaluation - with a big emphasis on process and policy learning.

Documentation: Peer-reviewed journal papers available

5. Home and Mobile Health Monitoring (HMHM)

Type: an experiment(?), as well as a policy demonstrator and learning pilot for telehealth and telecare

Duration: 2015-2018

Description: HMHM is a workstream within Scotland's Technology Enabled Care programme. The 3 year period from 2015 to 2018 can be presented either as a discrete programme or as a first phase of a longer term programme - the aim being in this first phase to "move from small scale pilots to large scale deployments informing scale-up and spread".

HMHM works with a group of 12 partners - a mix of NHS Boards and Health & Social Care Partnerships which competed to secure funding for participation in the programme. Partners could determine locally what population groups would be offered HMHM, and also - locally - identify through user engagement which technologies would be adopted. This is not the same as the user-

centred product design process used in Dallas. The partners were to deploy existing technologies, but they would consult potential users before making procurement decisions.

Evaluation methods: Although the evaluation report talks about 'impact', HMHM was not designed as an RCT. Nor did it attempt to include controls in other ways.

Documentation: Published report available.

6. United4Health

Type: EU research programme - an experiment as well as a policy learning pilot; telehealth.

Duration: The project ran for 3 years from 2013 to 2016.

Description: The full name of United4Health is **Uni**versal solutions in **Te**lemedicine **D**eployment for European **Health** care. It was funded though the ICT-PSP Competitiveness and Innovation Framework Programme. It is included here because Wales and Scotland were 2 of the 10 countries that participated, and. The 'ambition' was "to transform the patient experience through telehealth in Europe".

United4Health was similar to WSD in at least two key respects. It was conceived as a multi-site research programme (i.e. an experiment) as well as a policy learning pilot, and the number of participants reflected this – relatively large for a trial of this kind (approx. 7000), and relatively small in terms of the ambitions for *dallas* or *HealthSpace*.

Target groups: the selection criteria were similar to those for the telehealth arm of the WSD: people living in the community with a specified set of long-term conditions: COPD or heart failure or diabetes (though they also included hypertension).

Evaluation methods: Like WSD, the design called for the participation of people acting as controls (receiving 'usual care') to compare with the people receiving the intervention.

Documentation: Published report and peer-reviewed journal papers available for some of the individual demonstration sites.

7. NHS Test Beds programme

Duration: 1st wave launched in 2016; runs to 2018. 2nd wave launched autumn 2108.

Description: The Test Bed programme aims to "provide a *novel* (my italics) approach to test innovations in health systems, by tackling well established barriers to the uptake of innovation":

The 7 projects in the 1st wave (5 health and care Test Beds and 2 'Internet of Things' Test Beds) all run for 2 years and are being evaluated by outside institutions. All the 1st wave Test Beds involve testing ways of embedding or combining digital technologies into clinical pathways. They are "identifying solutions to support the transformation of clinical areas highlighted in the Five Year

Forward View including: primary care, urgent and emergency care, mental health, dementia and long term conditions". Three types of innovation are being tested in the programme:

- Predictive algorithms to manage patients at risk of developing a condition (risk stratification tools);
- Aggregation of data into one place to inform operational and clinical decision making and improve an individual's ability to manage their condition (shared / integrated records);
- Technology to monitor risk of crisis in clinical pathways or in individuals at home or care home (telehealth and telecare).

Evaluation methods: Each individual Test Bed takes a different approach to evaluation. Hence the methods being used are varied and include a randomised control trial, matched control studies, observational studies, feasibility studies and soft system methods.

Documentation: At the time of writing there are no published evaluations, though the programme has published a 'handbook' on evaluation which summarises lessons learned about evaluation from the 1st wave projects, and a brief report with a commentary on the process so far.

8. Social Care Digital Innovation Programme

Duration: The SCDIP started in 2017 and is now taking applications for the 3rd wave of funding.

Funding: Compared to most of the demonstrators included here, this is relatively small-scale (<£1 million in funding in wave 1). It is funded by NHS Digital in conjunction with the Local Government Association.

Description: "Over the last two years NHS Digital has funded 31 councils to provide digital solutions to social care challenges. In 2017-18 we saw Amazon Alexa introduced in Hampshire to reduce isolation and enhance independence; Luton and Central Bedfordshire improved the digital maturity of care homes and Norfolk increased referrals from customer services to appropriate voluntary sector support. In 2018-19 Nottingham is exploring an outcomes portal, Isle of Wight is testing exoskeletal devices for informal carers and Wirral is investing in biometric wearables to better understand support for people with autism"[II]. All projects work with existing technologies. The current round is looking for examples of innovative uses of digital technology in service design to improve the service user's experience and/or enable better service planning and commissioning.

Evaluation methods: Individual projects are being evaluated separately.

[II] https://www.local.gov.uk/our-support/our-improvement-offer/care-and-health-improvement/informatics/local-investment-programme

11. National demonstration programmes with a service/public health focus

9. Healthy Communities Challenge Fund/Healthy Towns Programme

Type: conceived as a policy experiment but not implemented as such?

Duration: 2008-11

Cost: £30 million

The Healthy Community Challenge Fund (HCCF) was set up to stimulate a 'whole town' approach to addressing the environmental determinants of obesity by testing and validating holistic approaches to promoting physical activity *etc*. that implement the lessons of existing programmes combined with efforts at galvanising members of the local community to take action to change both food and activity habits. As with HealthSpace, the government had a particular model in mind, in this case the EPODE1 model from France.

Local authorities and Primary Care Trusts were invited to bid jointly for funding, with a limit of £5 million per town, which had to be match-funded from local sources. The bidding process involved the submission of a 'coherent cross-sectoral plan that implemented a programme of interventions in the local area. In the event, nine towns and urban were selected, and they set up a total of 306 separate projects between them.

Target groups: the general population

Evaluation methods: Separate national and local evaluations, with the national evaluation - essentially a process evaluation - concentrating on common issues, e.g. How was the 'whole town approach' brief translated into practice? How was evidence used in the design of interventions? What were the barriers and facilitators to the implementation of a complex systems approach to obesity prevention?

Documentation: Published report for the national evaluation and peer-reviewed journal papers.

10. Partnerships for Older People Projects (POPP)

Duration: 2006-09

Estimated cost: £60 million

Description: The programme was funded by the Department of Health to develop services for older people that would help promoting their health, well-being and independence; and prevent or delay their need for higher intensity or institutional care. Twenty-nine local authorities were selected as pilot sites, all working in partnerships with health and voluntary sector partners.

In total, the 29 sites set up 146 core local projects, comprising many more individual services, aimed at improving health and well-being among older people and reducing social exclusion and isolation. The individual projects were determined according to local priorities. Of the 146 projects, two-thirds were primarily directed at reducing social isolation and exclusion or promoting healthy living among older people ('community facing'). The remaining one-third focused primarily on avoiding hospital admission or facilitating early discharge from acute or institutional care ('hospital facing'). Some addressed the full spectrum of needs. In addition to these 'core' projects, a further 530 small 'upstream' projects were commissioned from the third sector.

Target groups: Older people

11. NHS Diabetes Prevention Programme

The aim of the NHS DPP is "to deliver at scale, evidence based behavioural interventions for individuals identified as being at high risk of developing Type 2 diabetes". The DPP is not a trial. It is a national programme which is being rolled-out in waves after an initial 'demonstrator pilot'. Seven 'demonstrator sites' from across England were selected in 2015 to provide a variety of T2D prevention programme models, offering examples of intervention service delivery to inform subsequent NHS DPP development and roll-out (although all the 1st wave interventions rely on face-to-face programme delivery). Individuals are referred to programme by primary care physicians, which means that the programme has to identify individuals at risk of T2D and invite them to enrol. The first wave of the NHS DPP was then commissioned nationally, to be implemented by a national and regional team, and delivered by four provider organisations that had capacity to deliver the intervention across England.

Why pilot the DPP at all? Although it seems not be spelt out in the available documentation and reports, the key point seems to be that it requires system change (innovation in case-finding and data collection as well as commissioning of new services) and relies for success on uptake by members of the public who do not have a diagnosed disease: for the prevention programme to work *at scale*, large numbers of people (the target is 100,000 by 2020) have to enrol on the programme and adhere to it. In 2017 it was reported that the programme was making "good progress", with just

under 50,000 people referred in Wave 1 and more than 18,000 on the programme at the end of April. This exceeds the original target set in the NHS Mandate of 10,000 referrals during 2017/18[mm].

There are now resources available online which summarise learning from demonstrators, and these include (i) a 'readiness' checklist (ii) 10 top tips for rolling out the NHS DPP etc.

Independent formative evaluation of (i) the demonstrator site phase and (ii) first wave implementation was commissioned to inform subsequent NHS DPP implementation and evaluation. The evaluators have provided detailed and explicit recommendations based to inform the national programme management group and other decision makers on programme quality, improvements and future implementation and evaluation.

The programme now has a digital workstream to take advantage of the opportunities offered by digital technologies "to improve patient experience, and deliver services in a more efficient way reducing the burden on clinicians and the public". The upshot is that one of the NHS Test Beds is funded through NHS DPP - as part of digital workstream.

[mm] https://www.educationforhealth.org/news/wave-2-healthier-nhs-diabetes-prevention-programme-launches/

Part 4 References

Alexander H (2018). Towards scaling up Home and Mobile Health Monitoring: an evaluation of the outcomes achieved by year 3 and progress towards scale-up, spread and sustainability. TEC Scotland.

Chrysanthaki, T *et al* (2013). Stimulating whole system redesign: Lessons from an organizational analysis of the Whole System Demonstrator programme. *Journal of health services research & policy,* **18**(1_suppl), pp. 47-55.

Cummins, S *et al* (2016) National Evaluation of the Healthy Communities Challenge Fund: The Healthy Towns Programme in England. Technical Report.

Devlin AM *et al* (2015). Delivering digital health and well-being at scale: lessons learned during the implementation of the dallas program in the United Kingdom. *Journal of the American Medical Informatics Association,* **23**(1), pp. 48-59.

Elzen B *et al* (1996). Socio-technical networks: how a technology studies approach may help solve problems related to technical change. *Social Studies of Science*; 26:95-141.

Ettelt S *et al* (2015). The multiple purposes of policy piloting and their consequences: three examples from national health and social care policy in England. *Journal of Social Policy*; 44(2): 3109-337.

Galea A *et al* (2017). NHS Test Beds: the story so far. NHS England.

Giordano, R. *et al* (2011). Perspectives on telehealth and telecare: learning from the 12 Whole System Demonstrator Action Network (WSDAN) sites. Kings Fund.

Goodwin DM *et al* (2013). The role and status of evidence and innovation in the healthy towns programme in England: a qualitative stakeholder interview study. *J Epidemiology Community Health*; 67: 106-112.

Greenhalgh T *et al* (2010). Adoption, non-adoption, and abandonment of a personal electronic health record: case study of HealthSpace. *BMJ (Clinical research ed.),* 341, pp. c5814. See also The devil's in the detail: final report of the independent evaluation of the Summary Care Record and HealthSpace programmes.

Greenhalgh T, Russell J, (2010). Why do evaluations of eHealth fail? *PLoS Medicine*; 7(11): e1000360.

Greenhalgh T *et al* (2017). Beyond adoption: a new framework for theorizing and evaluating Non-adoption, Abandonment, and challenges to the Scale-up, Spread and Sustainability of health care technologies. *J. Medical Internet Research*; 19(11): e367

Henderson, C. *et al*, 2013. Cost effectiveness of telehealth for patients with long term conditions (Whole Systems Demonstrator telehealth questionnaire study): nested economic evaluation in a pragmatic, cluster randomised controlled trial. *BMJ (Clinical research ed.),* 346, pp. f1035.

Lennon MR *et al* (2017). Readiness for Delivering Digital Health at Scale: Lessons From a Longitudinal Qualitative Evaluation of a National Digital Health Innovation Program in the United Kingdom. *Journal of medical Internet research,* **19**(2), pp. e42

May CR et al (2009). Development of a theory of implementation and integration: Normalization Process Theory. *Implementation Science*; 4:29.

Moore G *et al* (2015). Process evaluation of complex intervention: Medical Research Council Guidance. *BMJ*; 350:h1258

MRC (2006). Developing and evaluating complex interventions. Medical Research Council.

Somers C *et al* (2019). Valuing mobile health: An open-ended contingent valuation survey of a national digital health program. *JMIR mHealth and uHealth,* **7**(1).

Rasmussen J. (2016). Universal Solutions in Telemedicine Deployment for European Health care. Final Report to ICT Competitive and Innovation Programme.

Windle K *et al* (2009). National Evaluation of Partnerships for Older People Projects. Final Report. Personal and Social Services Research Unit.

www.ageing.ox.ac.uk