



EUROPE

Better evaluations to support the needs of older people in the UK

Tom Ling, Alex Sutherland, Jack Pollard

For more information on this publication, visit www.rand.org/t/RR2932

Published by the RAND Corporation, Santa Monica, Calif., and Cambridge, UK

© Copyright 2019 RAND Corporation

RAND® is a registered trademark.

RAND Europe is a not-for-profit organisation whose mission is to help improve policy and decision making through research and analysis. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

Limited Print and Electronic Distribution Rights

This document and trademark(s) contained herein are protected by law. This representation of RAND intellectual property is provided for noncommercial use only. Unauthorized posting of this publication online is prohibited. Permission is given to duplicate this document for personal use only, as long as it is unaltered and complete. Permission is required from RAND to reproduce, or reuse in another form, any of its research documents for commercial use. For information on reprint and linking permissions, please visit www.rand.org/pubs/permissions.

Support RAND

Make a tax-deductible charitable contribution at
www.rand.org/giving/contribute

www.rand.org

www.rand.org/randeuropa

Preface

This report addresses the problem of how to improve the evaluation and evaluability of services for older people. This project followed a previous RAND report for Age UK on the evaluation of the Integrated Care Programme (ICP) (Sutherland et al. unpublished), where RAND's task was to review and make suggestions for how to improve the evaluation of these complex interventions. That work highlighted there were wider problems with how evaluation of services for older people were being conducted.

The present report describes the results of a series of three workshops with different stakeholder groups. These groups convened around the idea of improving how evaluations of initiatives in older care are commissioned, undertaken, reported and utilised.

This is a report funded by Age UK, one of the largest providers for the care of older people in the UK.

Dr. Tom Ling

Dr. Alex Sutherland

Jack Pollard

For more information about this report, please contact:

Dr. Tom Ling

RAND Europe

Westbrook Centre, Milton Road,

Cambridge, CB4 1YG

Email: tling@rand.org

Table of contents

Preface	iii
Figures	vi
Tables	vii
Executive summary	viii
Acknowledgements	ix
1. What is the problem?	1
1.1. Unfulfilled promise	1
1.2. Scope and structure of this concept note	2
2. What should be done differently?	5
2.1. Introduction	5
2.2. Building a community of practice to strengthen knowledge-sharing about approaches, methods and findings	5
2.3. Problem-focused evaluation and the co-production of evaluative knowledge	7
2.4. New models of commissioning and funding evaluation	9
2.5. Managing the tension between local and organisational needs and generalisable knowledge – better monitoring and routine data that can be compared	11
2.6. Knowledge sharing, rebalancing competition and cooperation, creating a ‘commons’ for shared knowledge and avoiding the ‘tragedy of the commons’	12
3. Conclusions and recommendations	14
3.1. Limitations and strengths	14
3.2. Discussion	14
3.3. Recommendations	14
References	17
Annex A. Responses from the warm-up question at each workshop	19
A.1. Evaluator workshop	19
A.2. Commissioner workshop	19
A.3. Provider workshop	20

Figures

Figure 1: Differences between Major and Meso evaluations (Ling & Wang 2018) 8

Tables

Table 1: Workshop questions by workshop 2

Executive summary

Transforming health and care for older people is complex and demanding. Evaluating such efforts requires a range of approaches and ideas, and a reflection on ways to improve how evaluations are commissioned, completed and used in a changing policy landscape. This report by researchers at RAND Europe summarises three workshops undertaken with evaluators, commissioners of evaluations and services, and those delivering services that are evaluated with the explicit aim of addressing these questions. It was broadly agreed that the areas to address were:

- **Relevance:** evaluations were too often not focused on major challenges and consequently findings were ‘unsurprising’ or even trivial.
- **Timeliness:** evaluation findings were often not available when decisions had to be made.
- **Replication and lack of cumulative building of knowledge:** evaluations were often designed as stand-alone pieces rather than building on previous evaluations and contributing to future evaluations.
- **Reluctance to share knowledge:** competition, a lack of clarity in communication, and poor knowledge management/mobilisation led to a reluctance to share information, pool data, or drew on the evaluations of potential competitors.
- **Efficiency:** evaluations often failed to use routine data and ignored the costs imposed on service providers and users.

Following from this, the workshops generated key recommendations:

For commissioners of evaluations: Commissioners from multiple organisations should collaborate routinely to identify key cross-cutting challenges, jointly fund evaluations addressing common concerns, engage the evaluation community earlier when developing an evaluation specification and ensure budgets and timelines match the scale of the question(s) asked.

For evaluators: Evaluators should engage earlier with commissioners and better balance competitive with collaborative behaviours to inform commissioners, develop standardised data sets, and make better use of routine data.

For service providers: Service providers can be a great source of tacit knowledge, but they often experience evaluations as something to be endured rather than as something to shape. Providers should work with evaluators to develop better routine monitoring and data collection.

For national policymakers in NGOs and government: Support the conditions under which the ‘tragedy of the commons’ can be avoided by: identifying and communicating priorities for evaluation research locally and nationally; monitoring the evaluation landscape; encourage and reward organisations for engaging with national priorities; and help nourish the ‘commons’ through sharing and collaborating while also ensuring adequate competition and variety.

Acknowledgements

The research team would like to thank all workshop participants for their enthusiasm and positive engagement during all three workshops, Age UK for their commissioning of the research and their continual support throughout, and Rob Prideaux and Natasha Elmore for their quality assurance of an earlier version of this report.

1. What is the problem?

1.1. Unfulfilled promise

In the field of evaluation, there is much to be said for Picciotto's classic statement that 'evaluation can yield benefits well in excess of its direct costs... through wholesome effects on the productivity of expenditures and the responsiveness of the public organizations' (Picciotto 1999, 15). However, this value is only realised under certain circumstances. Michael Quinn Patton (2008) has emphasised that, if they are to add value, evaluations should be much more closely focused on the needs of decision makers. Ling (2012) has emphasised the importance of conducting evaluations in real time, while the intervention itself is unfolding. Meanwhile, Davies (2013) has emphasised that too often evaluations are attempted when the essential characteristics of evaluability are absent, weakening or completely undermining any findings. It is these challenges and ideas that underpin this report, commissioned by Age UK.

The challenges outlined above are generic to delivering value through evaluation but they are made no easier by the turbulent landscape facing those seeking to improve services for older people in the UK. Policies on ageing are multiple, potentially overlapping, complementary and/or conflicting, and constantly shifting.¹ They include housing, pensions, moving people out of acute settings and into community settings, addressing currently unmet needs, dealing with inequality and poverty in older age, social services, lifestyle choices and behaviour, care home provision, independence and person-centred work, self-care, and unpaid care and volunteering. Running through all of these factors is the challenge of balancing quality of care and cost considerations and managing the trade-offs either with or on behalf of older people (Imison et al. 2017).

Adding to the complexity is a high turnover of policy initiatives in the UK. This has taken two forms. The first is incremental changes to existing policies, and the second is the development of new models of care. The latter is particularly intended to respond to the growing numbers of older people with complex health and social needs, and especially frail older people with limited income, for whom current models are unsustainable.

Consequently, however skilled the evaluators might be, the needs of decision makers may be so confused and uncertain that interventions may change considerably before attributable results could be ascertained, or interventions may overlap with other actions that evaluating impacts and learning lessons becomes a fool's errand. The 'wholesome effects' that Picciotto (1999) hoped for are therefore often more notable by their absence. In this briefing we explore these challenges further and identify ways in which they might be overcome.

¹ For a summary of recent policies on ageing, see Centre for Policy on Ageing (2018).

1.2. Scope and structure of this concept note

This note summarises results from a series of workshops undertaken to better understand barriers to a more cohesive evaluation approach for older people’s services. The genesis of this project was Age UK’s Integrated Care Programme (ICP), which RAND Europe has previously undertaken work on for Age UK (Sutherland et al. unpublished). That work highlighted gaps between local and national evaluations of ICP, and wider problems with the commissioning of evaluations in the elderly care sector. As such, the workshops were intended to be much broader in scope than ICP, encompassing general principles about the commissioning and execution of evaluations of services for older people.

The workshops were structured around a series of questions, set out below in Table 1. Participants were asked to discuss these as a group or in pairs, with the workshop led by a facilitator or facilitators (Dr. Ling and Dr. Sutherland).

Table 1: Workshop questions by workshop

Evaluators’ Workshop
1. How can we improve the commissioning of evaluations? What should be done differently?
2. Can we change the way in which the knowledge from evaluations is collated by commissioners?
3. Are there are a common set of principles that might help with the previous questions? What might they look like?
Commissioners’ Workshop
1. How can we improve the commissioning of evaluations to better inform service commissioning? What should be done differently?
2. Can we change the way in which the knowledge from evaluations is collated and used by commissioners?
3. Are there are a common set of principles that might help with the previous questions? What might they look like?
Provider Workshop
1. How can we improve the commissioning of evaluations to better inform service providers? What should be done differently?
2. Can we change the way in which the knowledge from evaluations is collated and used in delivering better services?
3. Are there are a common set of principles that might help with the previous questions? What might they look like?

In the report that follows, we set out the themes that emerged from the workshops and include a response to those themes in the form of a commentary. The purpose of the commentary is to reflect on findings

from the workshops, as well as possible solutions. In places, this means discussing limitations to a given suggestion or thinking further about a possible suggestion.

We conclude with recommendations for several groups based on the workshops.

2. What should be done differently?

2.1. Introduction

In order to understand more deeply the nature of the problem and to identify potential improvements, RAND Europe and Age UK ran three workshops between December 2017 and April 2018. These involved meeting with evaluators, commissioners of evaluations and services, and those delivering services that are evaluated. To avoid repetition in the following sections, we will refer to these three workshops as Evaluators' Workshop, Commissioners' Workshop, and Provider Workshop. Workshops lasted for three hours. From these a clear set of themes were established and, based on the discussions in the workshops, there was broad agreement about the areas to be addressed:

- **Relevance:** evaluations were too often not focused on major challenges and consequently findings were 'unsurprising' or even trivial.
- **Timeliness:** evaluation findings were often not available when decisions had to be made.
- **Replication and lack of cumulative building of knowledge:** evaluations were often designed as a stand-alone piece rather than building on previous evaluations and contributing to future evaluations.
- **Reluctance to share knowledge:** competition, a lack of clarity in communication and poor knowledge management/mobilisation led to a reluctance to share information, pool data or draw on the evaluations of potential competitors.
- **Efficiency:** evaluations often failed to use routine data and ignored the costs imposed on service providers and users.

In response to these challenges, five improvements were identified. We look at each of these in turn in the following sections.²

2.2. Building a community of practice to strengthen knowledge-sharing about approaches, methods and findings

2.2.1. What the workshop groups told us

The idea of a group that could work across organisational boundaries to share knowledge, identify shared problems and work together to address these problems lies at the heart of what Etienne Wenger and others have conceptualised as a 'community of practice' (Lave & Wenger 1991; Wenger et al. 2002). The Evaluators' Workshop noted that there had been some success in coordinating efforts (the Charity Evaluation Working Group (ChEW) and the convening work by New Philanthropy Capital were both mentioned), but despite these efforts the problems identified in the previous section had remained largely unchanged. More recently, the work of The Healthcare Improvement Studies (THIS) Institute is an

² Workshop participants were also asked 'warm up' questions about who loses out when evaluations are not well-commissioned. The responses to that question from each workshop are included in Annex A.

attempt to coordinate improvement research across the UK and might provide a basis for collaboration. THIS Institute is also using crowdsourcing to provide data and evidence for evaluations, which could potentially form part of a ‘commons’ available for all (evaluators, commissioners and providers) to draw upon. The Evaluators’ Workshop also thought that evaluators should be more proactive in seeking opportunities to collaborate, perhaps with more subcontracting to allow each group to focus on their areas of strength.

The Commissioners’ Workshop identified significant opportunities for commissioners to develop more collaborative ways of working with not only evaluators but also service providers and service users. Working in this way, and with each other, commissioners would become ‘better buyers’ of evaluation. At the same time, the Provider Workshop also emphasised that they would welcome a more inclusive approach to determine which questions to prioritise and to achieve the benefits of co-production. They thought that the Red Cross, Alzheimer’s UK and Dementia UK all had examples of good practice in sharing learning. However, in general it was felt that ‘we always send off data... but never see the results’.

2.2.2. Commentary

Communities of Practice (CoPs) do not offer a solution to all problems. However, they are considered to have ‘home’ areas where their strengths are most apparent. In the UK, as elsewhere, the particular strengths of CoPs in health and care have been linked to learning, sharing knowledge, and improved working together:

1. Strengthening learning and reinforcing the capacity to act on lessons learned (Nicolini et al. 2016).
2. Facilitating multi-professional information and knowledge sharing (Richardson 2016).
3. Contributing to a more general need to achieve better ways of working, to deliver high-quality care within constrained budgets, by improving productivity (Ranmuthugala et al. 2011).

The challenge to creating a culture of learning that can improve quality and safety in health and social care is raised throughout the Francis Report on the Mid Staffordshire NHS Foundation Trust (Francis 2013) and these themes are returned to again in the Berwick Report on patient safety (Berwick 2013). The barriers created by disciplinary and organisational boundaries and competing incentives can be, at least in part, addressed by CoPs (Nicolini et al. 2016). All three workshops highlighted the potential benefits of improved learning from evaluations and, while CoPs were not always referred to by name, the desirability of working in this way was often referred to. More widely, we are seeing more confidence in the ability of CoPs to support knowledge sharing and collective learning (Wenger 2010) and improving efficiency and productivity (Ranmuthugala et al. 2011). However, a note of caution should also be expressed because CoPs themselves have not been systematically evaluated. This is especially true for an innovative CoP that might be developed for improving evaluation to benefit older people: ‘Although there is widespread acceptance that CoPs are useful, formally assessing their value and understanding what helps them be more effective is less easy: they are complex social systems, and it is inherently difficult to directly connect cause and effect’ (Van Winkelen 2016, 3–4).

2.3. Problem-focused evaluation and the co-production of evaluative knowledge

2.3.1. What the workshop groups told us

The Evaluators' Workshop was very clear that stronger co-working between evaluators and commissioners would be beneficial. Having commissioners and evaluators engaging early and throughout the evaluation process (for example, through precompetitive workshops to help frame evaluations) and greater use of advisors when developing the invitation to tender (ITT) (and possibly sitting alongside commissioners during implementation) were both seen to be a practical means to strengthen these links. A comparison with the construction industry, where such arrangements are apparently common, was made. Equally, the Evaluators' Workshop felt that evaluators need to be more collaborative and self-aware for this to work. At a practical level, higher standards of project management among evaluators would also make engagement easier.

At the same time, the Evaluators' Workshop saw that delivery partners should be more involved in this process and in particular to help identify how the wider system might be altered by an intervention. Delivery partners in smaller organisations should be encouraged to work together to build larger, more meaningful and better-powered evaluations. In this respect, there might be efficiencies in setting up standard data sets rather than collecting different categories of information for each evaluation. Regardless of the changes proposed, it was emphasised by the Evaluators' Workshop that including the service user voice is important (although concrete examples of this being done well were hard to find). We note that healthcare evaluators funded by, for example, NIHR, require Public Patient Involvement (PPI), which may be a model to mirror in social care evaluation.

Commissioners' views in a sense reflected the other side of this coin by emphasising the potential benefits from evaluative evidence being used in helping to design and set up services and supporting continuous service improvement. The Commissioners' Workshop also recognised that commissioners need to be clearer about what information and analysis they want from the evaluation. Commissioners reflected the view that there would be a benefit in understanding the whole system rather than having narrowly focused evaluation questions (although these might also be needed in some cases). However, commissioners also stressed the importance of co-designing tenders with frontline staff and service users (while recognising that this can be difficult). Commissioners were especially anxious about how they could make sense of and use evaluation products and were clear that long technical reports prevented them from taking ownership of findings.

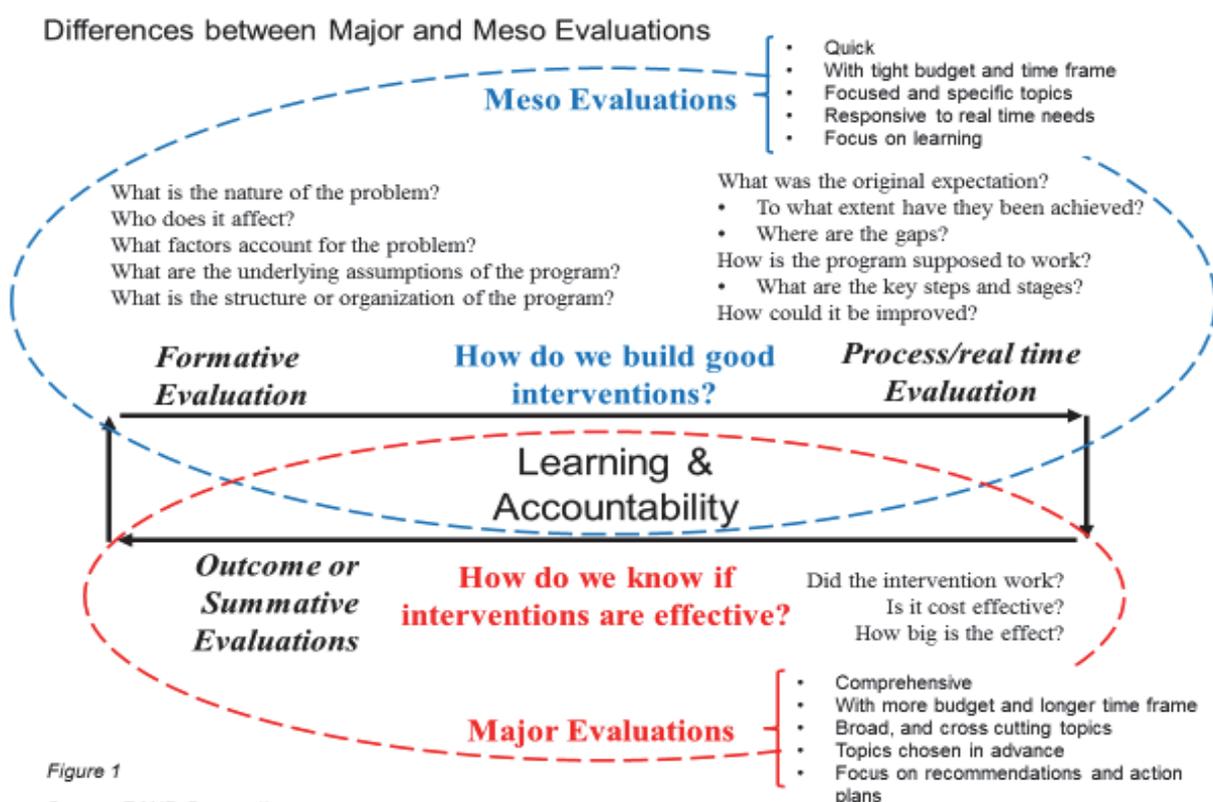
The Provider Workshop emphasised the significant problems caused by not being directly involved in the co-production of knowledge, and experiencing evaluation as something that is done to them rather than done with them. Co-production and standardisation of evidence was seen to be especially helpful and the case of Dementia UK having an overarching framework that was replicable across the UK was mentioned. The Provider Workshop also emphasised the potential tension between focusing on system wide problems and providing them with evaluations that could help them deliver their particular services better: 'We ended up doing our own evaluation because some of our questions weren't being answered.' However,

despite emphasising the ‘competitive-interdependence’ that exists among providers, they were optimistic about the possibility of effectively sharing insights and learning across organisations.

2.3.2. Commentary

It is worth distinguishing between different evaluation processes and products. Smaller evaluations might be conceptualised differently to major impact evaluations. However, in all cases, the benefit of achieving a co-production of knowledge that can be readily socialised among decision makers, practitioners and service users is apparent. In particular, many evaluations might be conducted in so-called ‘real time’ where those implementing the programme are still learning and adapting in the light of emerging evidence (Ling 2012). This distinguishes real-time evaluations from ‘ex post’ evaluations that seek to form a judgement about something that is complete. In this situation, the knowledge being created is often uncodified and tacit and may well be fluid as both providers and service users reflect upon and reconsider the services provided. Embedded evaluation can play a helpful role in exploring and sometimes challenging these assumptions. In a separate study for the World Bank, Ling and colleagues conceptualised this relationship between real-time and major evaluations and learning in the following way:

Figure 1: Differences between Major and Meso evaluations (Ling & Wang 2018)



One topic of discussion was the idea of ‘co-production’, which is very much ‘in vogue’ in discussions of evaluation. ‘Co-production’ as a concept dates from the 1970s and was introduced by Elinor Ostrom (1996) to describe a situation where inputs from ‘individuals who are not “in” the same organization are

transformed into goods and services' (Ostrom 1996, 1073). This included both service users and providers in making contributions. More recently, it has become a widespread term used in discussions about public policy in general, and health and social care in particular (Carr 2016; Needham 2008). However, we agree with Filipe and colleagues that the agreement that it is a 'good thing' is greater than the understanding of what it may mean in practice (Filipe et al. 2017). They go on to argue:

One way of going about the co-production of health care more meaningfully is to look at it as a dynamic, experimental, and reflective process sustained by different forms of engagement, interactions, and social relations and that may generate, in turn, **new forms of care other than health care** (e.g., inclusive relationships, solidarity), **values beyond economic value** (e.g., equity, justice), and **new insights and research practices** that are relevant to different disciplines and practices (e.g., community participation, patient advocacy, collaborative research). (Filipe et al. 2017, 5).

So co-production, which was supported by our workshop participants, may be a very helpful way of thinking about how to conceptualise the different contributions of stakeholders, but it should not be seen as a simple method that can guarantee their full engagement in knowledge production.

2.4. New models of commissioning and funding evaluation

2.4.1. What the workshop groups told us

The Provider Workshop identified significant problems with the existing models of commissioning and funding evaluation. Participants felt these often led to evaluations that were disproportionate:

'The evaluation becomes more important than the thing you were trying to do in the first place. You end up trying to service the evaluation rather than helping the people you want to do.'
(Participant: Provider Workshop).

Existing models of commissioning and funding were said to focus on questions that were not relevant. Commissioners also failed to make use of routine data resulting in duplicated effort, and they failed to share results. The Evaluators' Workshop recognised similar problems and advocated a radical shift in the model for commissioning evaluations. Evaluators felt that commissioners typically work up an ITT, select a bidder primarily on price, and then have a few check-in steps focused on milestones to trigger payment. Instead, the Evaluators' Workshop suggested a top-to-toe change involving: the evaluation community taking more responsibility for identifying what was evaluable and resisting incentives to conduct evaluations that were unfeasible; more open engagement in developing terms of reference; more interactions between evaluators to identify the best approaches and best teams without losing the benefits of competition (competitive inter-dependence); more time taken in problem identification at the start of a contract; and clear 'step-back and adapt' moments during the life of the evaluation to ensure its continued relevance. Less specifically, the Evaluators' Workshop also emphasised the value of including service users' voices and experiences.

The Commissioners' Workshop articulated related principles, underpinned by an understanding that evaluations that show things do not work, or do not work as expected, are just as important as evaluations that find positive outcomes and therefore need to be shared in the same manner. However, it should be

collectively acknowledged that a further dialogue between researchers and those funding evaluations on how to present and discuss results is needed.

- Stakeholders with different perspectives should be incorporated in the commissioning stage, so it is fully understood what needs to be evaluated before the evaluation gets under way.
- There should be a development stage/approach when commissioning evaluations so that the scope of the potential evaluation is clearly defined.
- Evaluators should provide a skeleton report to commissioners while the evaluation is being undertaken, so that the commissioner can provide feedback on what they want in the final report.
- Evaluators should tailor the outputs of their evaluations to different audiences (e.g. slide deck, short report and infographics) to improve knowledge collation and accessibility.
- Lessons need to be learned from past mistakes, as well as past successes, when future evaluations are to take place.
- ‘Sand-pitting’ (similar to that of a market testing event, but less formal and more collaborative) could be used to help commissioners better understand how issues can be tackled/services evaluated.
- Throughout the implementation and running of a service, it would be useful to gather numerous stakeholders in an open environment and explore how different actions and scenarios could be undertaken and the potential outcomes that could be realised – for example, through a gaming exercise.
- Commissioners should communicate with each other to avoid duplication and create synergies.

2.4.2. Commentary

There are many good practice guides to commissioning evaluations.³ However, these are almost all technical in nature, providing a step-by-step guide to commissioning independent evaluations within the existing way of working. The workshop participants articulated an appetite for something more ambitious and more strategic. This interest is consistent with participants’ sense that the existing model of commissioning evaluations was compatible only with a particular, and relatively uncommon, set of circumstances, namely:

- The definition and boundaries of the thing being evaluated were clear and agreed
- The outcomes of interest were similarly agreed and defined
- The processes and activities were known and visible
- The interactions with other parts of the environment were understood and measurable.

While these circumstances do sometimes occur, in most cases of evaluating services for older people they do not.

However, unsurprisingly given the scale of the task and the time available, the workshops fell short of a clearly defined set of alternative processes. It is unlikely that any new approach to commissioning

³ Including, for example: Commission Accomplished (2018); McNeish & Scott (n.d.); MCN Evaluations (2018); United Kingdom Evaluation Society (2013).

evaluations would get it right first time. Therefore, an approach that first built on good inclusive design processes, such as those proposed by the Design Council (2018), and then maintained a watching brief over progress would be preferable at this stage to a wholesale change.

2.5. Managing the tension between local and organisational needs and generalisable knowledge – better monitoring and routine data that can be compared

2.5.1. What the workshop groups told us

The Provider Workshop participants reiterated the importance of evaluation products that focused especially on the local and specific service. They were worried that they were missing opportunities for service improvement and that the really ‘ugly’, ‘hidden interior’ of programmes might not be brought to light by the evaluation. They recognised that the gap between tailoring evaluations to local needs and contributing to wider evaluations might be unbridgeable and that therefore there would be a need for different kinds of evaluation. The Evaluators’ Workshop was in general less spontaneously concerned with this aspect, but there was a view that the work currently being trialled by THIS Institute on using crowdsourcing to identify issues from those ‘in the field’ was worth considering further. Both the Evaluators’ Workshop and the Commissioners’ Workshop were interested in how standardised data collection might help bridge the gap between the needs of local service providers and the accumulation of scientific knowledge over time and place.

2.5.2. Commentary

There are at least two tools that help to address this tension. First, it is possible to systematically collect and analyse data on the different contexts of local services. This helps to explain variations in outcome that are not caused by the service itself but by the context in which it is being implemented. By using Qualitative Comparative Analysis, it is possible to analyse the causal contribution of different conditions to an outcome of interest.⁴ This is especially useful when, as is common in services for older people, there may be many causal paths to achieving an impact, rather than just a single direct route, or there may be many routes to impact but one particular blocker that guarantees failure. By including more systematically collected contextual and programme specific evidence, it becomes easier to compare and contrast different programmes in different settings.

Second, there is a large and growing interest in not simply *describing* the processes and activities associated with a programme or service, but also seeking to identify the underlying mechanisms that generate the outcomes of interest. This is most commonly associated with realist evaluation,⁵ but there has been long-standing interest in understanding what brings about change (Pawson & Tilley 1997; Hedstrom & Swedberg 1998; Glennan 1996). This recognises that programmes are complex and are embedded in

⁴ For a comprehensive definition of Qualitative Comparative Analysis, see Davies (2016).

⁵ Realist evaluation considers not only the outcomes of an intervention but also what works, under which circumstances, and for whom. This was an approach first developed by Pawson & Tilley (1997).

different ways in different contexts, so that the important thing to understand is whether and how underlying mechanisms of change have been masked or amplified by local circumstances.

2.6. Knowledge sharing, rebalancing competition and cooperation, creating a 'commons' for shared knowledge and avoiding the 'tragedy of the commons'

2.6.1. What the workshop groups told us

Across all three workshops, there was an understanding that the organisational structures, incentives and accountabilities of services for older people could very often lead to the 'tragedy of the commons'; that is, where things of common interest are not achieved because no person or organisation feels responsible for nurturing and maintaining them, and/or the 'free rider' problem leaves everyone hoping that someone else will take the necessary action. In this case, a specific issue was that each organisation providing services for older people was, to a degree, competing with other organisations and that consequently there was a natural reluctance to share information.

Commissioners were also reluctant to share findings when they may have lost faith in previous evaluations or felt that the quality of their evidence was weak. Similarly, there were genuine problems associated with using the results from a specific and limited evaluation to generalise across a wider range of services and, appropriately, commissioners did not feel able to do this because the resulting knowledge may be very localised in nature. Evaluators noted that their colleagues had no incentive to limit the number of evaluations (and thus the amount of work commissioned) or share findings, and that contracts might often prevent sharing of findings. They looked to commissioners to find new models of commissioning that would create incentives to explicitly address shared and common interests – especially cross-cutting evaluations. A further point from the Evaluators' Workshop was that knowledge mobilisation should be considered to be a separate task and funded accordingly; once the evaluation was complete, evaluators typically would find themselves moving onto the next piece of work before knowledge had been mobilised. In this context, skills in sense-making and messaging for different audiences were also felt to be absent. In general, therefore, it was said that there was neither the incentive nor the capacity in the current system to foster a 'vibrant commons'.

2.6.2. Commentary

Garrett Hardin in 1968 asserted the tragedy of the commons in *Science* magazine, suggesting that individuals who serve the interests of the group will lose out to self-serving individuals (Hardin 1968). This bleak vision was compellingly challenged by the work of Nobel Prize-winning economist Elinor Ostrom (1990, 2010) who showed that groups can (and in practice often do) take action to preserve the commons, without the need for top-down regulation, providing that eight core design principles are met: 1) Clearly defined boundaries; 2) Proportional equivalence between benefits and costs; 3) Collective choice arrangements; 4) Monitoring; 5) Graduated sanctions; 6) Fast and fair conflict resolution; 7) Local autonomy; and 8) Appropriate relations with other tiers of rule-making authority (polycentric governance).

These principles provide a useful checklist to assess how far organisations providing services for older people could collaborate to maintain a commonly held body of knowledge. They might lead to the conclusion that the tragedy of the commons could be averted if:

1. It was clearly established what information would be shared and what would be confidential to each organisation.
2. Organisations committed to contribute broadly in line with their resources.
3. Priorities for evaluation research could be collectively set.
4. Progress could be monitored.
5. Organisations that failed to 'play the game' could be brought back into line through graduated sanctions.
6. Disagreement and disputes could be speedily resolved.
7. The need for organisational autonomy was respected.
8. Support for shared evaluation approaches was nested within wider accountabilities and responsibilities.

3. Conclusions and recommendations

3.1. Limitations and strengths

It is important to recognise the limitations of this work. The number of workshops and participants was small and the time available in each one was limited. However, the participants were all experts in their own fields and deeply immersed in the questions relevant to this report. Furthermore, each workshop tended to reinforce rather than undercut the messages from the other workshops. Additionally, comments from the workshop resonated with the wider literature. While facilitating the workshops, there was no sense that participants were ‘holding back’ information or otherwise dissembling.

3.2. Discussion

The workshops reinforce a wider sense that something is not right in how we commission, deliver and use evaluations, and that this is not the responsibility of individuals or groups but a consequence of the way the social activity of evaluation is organised. In relation to evaluating services for older people, we have seen the particular barriers that exist to improvement. The consequence is that we have evaluations that are suboptimal in terms of the questions being asked and the answers provided, but moreover that, at present, there is little opportunity to generalise and no systematic collation of knowledge generated by these evaluations. The solutions to these are as much social as they are technical; both are required to achieve improvement.

3.3. Recommendations

3.3.1. For commissioners of evaluations

- Revisit the standard model of commissioning evaluations with a view to finding a greater variety of approaches to support more collaborative efforts by the evaluation community. This could include pre-commissioning workshops with researchers/evaluators to understand the parameters of a given evaluation approach/topic.
- Reinforce expectations that evaluations should be based on thoroughly understood problems.
- Hold evaluators to account through more flexible means with scheduled ‘step-back and adapt’ moments throughout the evaluation.
- Ensure that service providers’ and users’ voices are more strongly present in the evaluation process.
- Collaborate with commissioners from other organisations to identify cross-cutting challenges and jointly fund evaluations driven by these shared problems.
- Ensure that the evaluation aims and timeline are matched by appropriate funding.
- Thinking about the overall evaluability of an initiative, ideally through involving evaluators early in the process, either through consultation with ‘critical friends’ or through early ‘market engagement’ prior to/during the development of an evaluation specification.

3.3.2. For evaluators

- Reach out to other evaluators to create a forum for identifying challenges and mobilising knowledge.
- Explore the use of standardised data and the more systematic collection of contextual and service/programme specific data, for example, through Qualitative Comparative Analysis, and inform commissioners about what is technically possible and helpful.
- Develop more innovative evaluation products tailored to the needs of different audiences.
- Explore the better use of routine data to minimise the burden of data collection and inform commissioners about how this might best be done.
- Improve the ways of working with service providers and ensure that results are relevant and shared.
- Work at how to understand and inform expectations of commissioners and present the results of research accessibly.

3.3.3. For service providers

- Engage with both commissioners and evaluators to make tacit and lived expertise more visible in the evaluation process and ensure that evaluations are focused on real problems.
- Ensure that they are monitoring services and collecting routine data appropriately to improve the overall evaluability of initiatives (see also section 5.7 in Sutherland et al. unpublished).

3.3.4. For national policymakers in NGOs and government

Support the conditions under which the ‘tragedy of the commons’ can be avoided, by:

- Establishing what information would be shared and what would be confidential to each organisation.
- Ensuring organisations contribute broadly in line with their resources.
- Identifying and communicating priorities for evaluation research.
- Monitoring progress of the new evaluation landscape.
- Identifying organisations that failed to engage with evaluation and have graded responses from ‘naming and shaming’ through to funding consequences.
- Ensuring disputes can be speedily resolved.
- Respecting variety of providers and organisational autonomy.
- Supporting shared evaluation approaches that are nested within wider accountabilities and responsibilities.

3.3.5. For all groups working in the sector

- Explore the use of standardised data and the more systematic collection of contextual and service-/programme-specific data.

References

- Berwick, D. 2013. *A Promise to Learn – A Commitment to Act: Improving the safety of patients in England*. London: Department of Health.
- Carr, S. 2016. ‘Can Co-production Really Transform UK Mental Health Services?’ *Mad in America: Science, Psychiatry and Social Justice*, 29 August 2016. As of 03 October 2018: <https://www.madinamerica.com/2016/08/can-co-production-really-transform-uk-mental-health-services/>
- Centre for Policy on Ageing. 2018. ‘Policies on Ageing.’ As of 03 October 2018: http://www.cpa.org.uk/cpa/policies_on_ageing.html
- Commission Accomplished (homepage). 2018. As of 03 October 2018: <http://www.commissionaccomplished.co.uk/>
- Davies, R. 2013. ‘Planning Evaluability Assessments: A Synthesis of the Literature with Recommendations.’ Cambridge: Department for International Development. As of 03 October 2018: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/248656/wp40-planning-eval-assessments.pdf
- Davies, R. 2016. ‘Qualitative Comparative Analysis.’ *BetterEvaluation*, 5 January 2016. As of 03 October 2018: https://www.betterevaluation.org/en/evaluation-options/qualitative_comparative_analysis
- Design Council. 2018. ‘The Design Process: What is the Double Diamond?’ As of 03 October 2018: <https://www.designcouncil.org.uk/news-opinion/design-process-what-double-diamond>
- Filipe, A., A. Renedo, & C. Marston. 2017. ‘The co-production of what? Knowledge, values, and social relations in health care.’ *PLoS Biology* 15(5): e2001403. <https://doi.org/10.1371/journal.pbio.2001403>
- Francis, R. 2013. *Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry*. London: The Stationery Office.
- Glennan, S. S. 1996. ‘Mechanisms and the Nature of Causation.’ *Erkenntnis* 44: 49–71.
- Hardin, G. 1968. ‘The Tragedy of the Commons.’ *Science*. 162(3859): 1243–8.
- Hedstrom, P. & R. Swedberg. 1998. *Social Mechanisms: An Analytical Approach to Social Theory*. Cambridge: Cambridge University Press.
- Imison C., N. Curry, H. Holder, S. Castle-Clarke, D. Nimmons, J. Appleby, R. Thorlby & S. Lombardo. 2017. *Shifting the balance of care: great expectations*. London: Nuffield Trust.
- Lave, J., & E. Wenger. 1991. *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.
- Ling, T. & B. Wang. 2018. *Forthcoming*.
- Ling, T. 2012. ‘Evaluating complex and unfolding interventions in real time.’ *Evaluation* 18(1): 75–87.
- MCN Evaluations (homepage). 2018. As of 03 October 2018: <http://mcnevaluation.co.uk/>

McNeish, D. & S. Scott. n.d. 'Project Evaluation – A Practical Guide: Part 3 Commissioning an independent evaluation'. DMSS. As of 03 October 2018:

<https://www.dmss.co.uk/pdfs/Part-3-Commissioning-an-independent-evaluation.pdf>

Needham, C. 2008. 'Realising the Potential of Co-production: Negotiating Improvements in Public Services.' *Social Policy and Society* 7(2): 221–31.

Nicolini, D., H. Scarbrough, & J. Gracheva. 2016. 'Communities of Practice and Situated Learning in Health Care.' In *The Oxford Handbook of Health Care Management*, edited by E. Ferlie, K. Montgomery, & A.R. Pedersen, 255–78. Oxford: Oxford University Press.

Ostrom, E. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press.

Ostrom, E. 1996. 'Crossing the great divide: Coproduction, synergy, and development.' *World Development* 24(6): 1073–87.

Ostrom, E. 2010. 'Polycentric Systems for Coping with Collective Action and Global Environmental Change.' *Global Environmental Change* 20(4): 550–7.

Patton, M.Q. 2008. *Utilization-Focused Evaluation*. 4th edition. London: Sage.

Pawson, R. & N. Tilley. 1997. *Realistic Evaluation*. London: Sage.

Picciotto, R. 1999. 'Towards an Economics of Evaluation'. *Evaluation* 5(1): 7–22.

Ranmuthugala, G., J.J. Plumb, F.C. Cunningham, A. Georgiou, J.I. Westbrook, & J. Braithwaite. 2011. 'How and Why Are Communities of Practice Established in the Healthcare Sector? A systematic review of the literature.' *BMC Health Services Research* 11(1): 273.

Richardson, L. 2016. 'Sharing knowledge: Performing co-production in collaborative artistic work.' *Environment and Planning A: Economy and Space* 48 (11): 2256–71.

Sutherland, A., K. Saunders, M. Sim, J. Exley, K. Kruihof, E. Pitchforth, & T. Ling. Unpublished. *Improving the evaluation of integrated care*. London: Age UK.

United Kingdom Evaluation Society. 2013. 'Guidelines for good practise in evaluation.' As of 03 October 2018:

https://www.evaluation.org.uk/images/ukesdocs/UKES_Guidelines_for_Good_Practice_January_2013.pdf

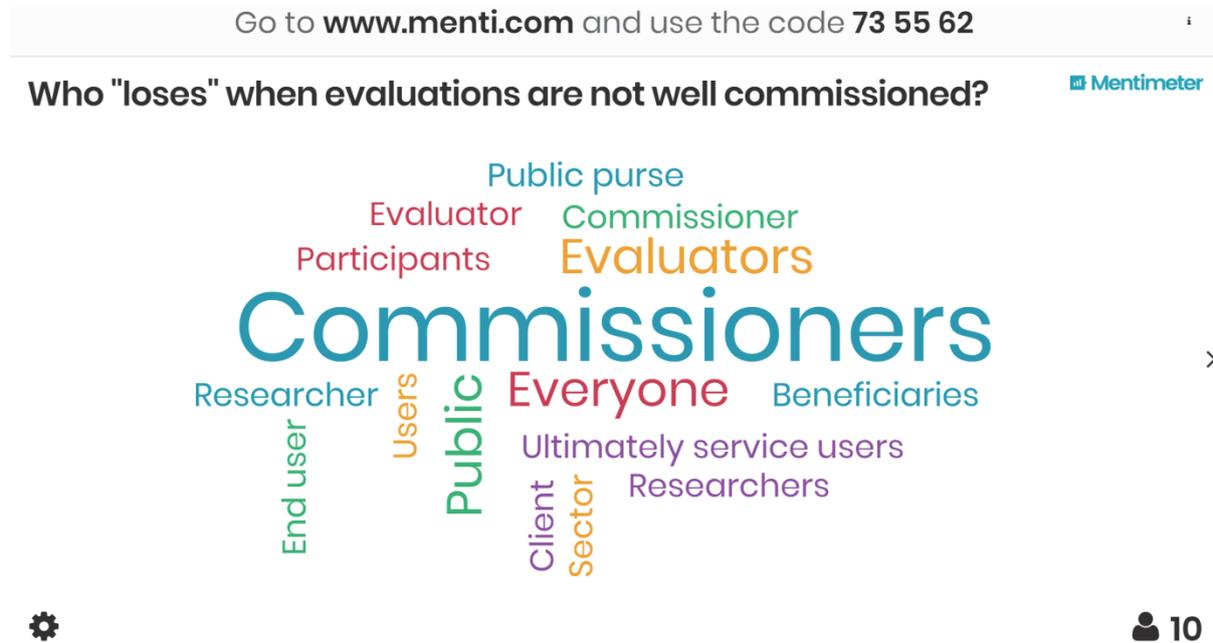
Van Winkelen, C. 2016. 'Using Developmental Evaluation Methods with Communities of Practice.' *The Learning Organization* 23(2/3): 141–55.

Wenger, E. 2010. 'Communities of Practice and social learning systems: The career of a concept.' In *Social Learning Systems and Communities of Practice*, edited by Blackmore, C., 179–98. London: Springer Verlag and the Open University.

Wenger, E., R.A. McDermott, & W. Snyder. 2002. *Cultivating communities of practice: A guide to managing knowledge*. Harvard: Harvard Business Press.

Annex A. Responses from the warm-up question at each workshop

A.1. Evaluator workshop



Although 'commissioners' was the most frequent response in the evaluation group, there was a high level of agreement overall that 'everyone' loses out when evaluations are not well-commissioned.

A.2. Commissioner workshop



A.3. Provider workshop

Go to www.menti.com and use the code **61 91 2**

Who "loses" when evaluations are not well commissioned? Mentimeter

The word cloud displays the following terms: Service user, Org who commissioned, Eveybody, Service users, Wider systems, Service users, Service staff, Providers, Users, Commissioners, Funders, Charity, Organisation, Service provider, Commissioners, Donors, Clients, World, Will, Funder. The word 'Everyone' is the largest and most prominent.

14