UNPICKING THE FABRIC OF METHODOLOGICAL CHALLENGE: EVALUATING ARTS-BASED ACTIVITIES FOR PEOPLE LIVING WITH DEMENTIA

K C R Gray

A thesis submitted in partial fulfilment of the University's requirements for the
Degree of Doctor of Philosophy

2020

University of Worcester

ABSTRACT

Background

As the human, social and economic costs associated with dementia become clear, it is hoped that the arts and culture will provide a valuable resource for meaningful and enjoyable activity that can help those affected by the condition to live well. Arts organisations, individual artists, museum educators, care staff, and members of the public are increasingly involved in delivering arts and cultural activities for those affected by dementia in their communities. To compete on equal terms for funding with pharmacological, clinical, or psychosocial treatments and interventions, convincing and auditable evidence for the benefits of the arts is required. This can be difficult to provide. To support good quality grass-roots provision, and to ensure that high-level public health commissioning is effective and well-targeted, it is crucial that the field as a whole understands how to better evaluate the work it is doing.

There has to date been no extended study of the methodological challenges facing evaluators across the sector. This research aimed to identify and describe the challenges as they are reported within the literature, and to explore how they are experienced by evaluation stakeholders.

Methods

This qualitative study included a large and systematically-informed narrative synthesis of the literature using a hermeneutic framework. This was followed by interviews with evaluators, researchers, artist practitioners and arts managers and commissioners. These were conducted and analysed using a critical realist informed approach to grounded theory. Initial theoretical categories identified in the first 21 interviews were tested and elaborated in five further interviews with highly experienced researchers and evaluators. A collaborative project with an artist resulted in the creation of an illustrated graphic narrative used to communicate and disseminate findings.

Results

Commonly experienced methodological difficulties are identified through the literature review, along with a set of sensitising concepts: value, context, ethics, meaning, and use. The metaphor of 'fabric' is introduced to describe methodological challenge and the way in which these concepts may be woven through it.

In interview, participants described anxieties and divisions experienced in evaluation when they had to make and act on decisions about what might be valuable or meaningful. They struggled to reconcile epistemic and non-epistemic values, particularly when trying to introduce standards of rigour to their practice. Interviewees' responses are discussed in the light of key concepts for the field of arts and dementia, including cultural value, evidence-based medicine, ethics and rights, and quality.

Conclusion

To help address the challenges identified for the sector as a whole, a novel values-informed approach to evaluation is proposed. This is based on the idea of arts and dementia activity as an 'ethical practice'. Foundations for this approach are suggested, comprising attention to ideas of complexity, multidisciplinary and collaborative working, methods innovation, and an acknowledgement of the problems posed by lack of evaluation capacity and resource. Recommendations are made for practice and future research to support exploration and testing of these ideas.

ACKNOWLEDGMENTS

This research was made possible by funding from the Alzheimer's Society for the TAnDem (The Arts and Dementia) Doctoral Training Centre. My sincere thanks go to all those involved, including Alzheimer's Society staff, the monitors, and colleagues at both the University of Nottingham and the University of Worcester.

I thank my research participants for their time, their honesty, and their wisdom, as well as for the work they do every day to the benefit of people living with dementia. I hope I have reflected their words and experiences accurately and that my research may prove useful to them in future.

The constructive and critical insight and the experience of my supervisory team – Simon Evans, Amanda Griffiths, and Justine Schneider – has been invaluable. They have calmly kept me on track. Mentors come in many forms: the ideas of Jane Willis, Norma Daykin, Louise Mansfield, and Clive Cazeaux have positively shaped my work. I have learned from many people. All my errors are, of course, my own.

Colleagues and friends at the Association for Dementia Studies created a working environment that grounded me, and often felt like coming home. It was wonderful to spend time working with Emma Lazenby of ForMed Films CIC. My fellow TAnDem PhD students have been a constant source of inspiration, encouragement, and fun: the link(ee)s and memories made over the last five years will be long-lasting.

Friends and family have supported me with bemusement, compassion, and good humour as I stumbled through this PhD thing for the second time. I know, it was crazy, and I blame you all for not stopping me! My parents, no longer with me physically, are always in my mind. My children, Jo and Laurie, are my harshest critics and my fiercest supporters: I love them to the moon and back, and only hope they can forgive me for all the hours spent at my desk over the past four and a half years.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

> Art […] is a means of union among men, joining them together in the same feelings, and indispensable for the life and progress toward well-being of individuals and of humanity.
>
> Leo Tolstoy, *What is art?*

1.1 Background to the arts and dementia

The considerable human, social, and economic costs of dementia are well documented (Pickett *et al.*, 2018). Not one disorder, but many, it affects heterogenous groups of individuals, each of whom may experience a variety of different symptoms. It is widely understood that there is currently no cure for the condition in any of its forms, and that those who have it will experience a progressive decline in cognitive functioning, including memory, often alongside a range of distressing physical and neuropsychiatric symptoms (Cerejeira, Lagarto and Mukaetova-Ladinska, 2012).

However, such a negatively deterministic view would rightly be challenged as problematic by anyone with direct experience of living with dementia or caring for someone living with the condition (Kitwood, 1997). A person with dementia is always considerably more than their diagnosis, ‘I am not simply a bundle of attributes’ (Bryden, 2020, p. 75). While few would wish to sugar-coat the often unpleasant realities of dementia, it is recognised that there are some things that can make it possible to live better, and sometimes even *well* with it.

Participation in activity that is meaningful and enjoyable to the individual is thought to improve the chances that a person with dementia will live well (National Institute for Health and Clinical Excellence, 2015). Factors identified in a recent review as being positively associated with a higher quality of life include greater social engagement and positive relationships with carers (Martyr *et al.*, 2018). Depression and neuropsychiatric symptoms, anxiety, pain and the presence of unmet needs, as well as carer burden or stress and distress at symptoms, were found to be negatively associated with quality of life in the same review. Nonpharmacological interventions that address the unmet needs of people with dementia are sorely needed; it is thought that responding to those needs can decrease the behaviours that often cause the most stress and distress to all involved (Cohen-Mansfield, 2018).

A case could be made for the importance and benefit of the arts for people with dementia on multiple grounds, including quality of life, psychology, economics and human rights (All Party Parliamentary Group on Arts Health and Wellbeing, 2017; Cousins, 2018). For people with dementia and their carers, viewing and making art together has been shown to help promote social inclusion and social engagement, enhancing the caring relationship (Camic, Tischler and Pearman, 2014). Being playful together in the context of a participatory arts programme offers potential for expression and creating connection (Swinnen and de Medeiros, 2018). Singing and music listening can improve mood, orientation, and memory in early dementia (Särkämö *et al.*, 2014). Activities that are mentally engaging, enjoyable, social, and which reduce stress have a role to play in improving resilience, even for those with advanced dementia (Newman *et al.*, 2019). They may even have a neuro-protective effect before a dementia is diagnosed (Fancourt, Steptoe and Cadar, 2018). The arts have also been instrumental in supporting the professional development of dementia care staff, improving communication and validating their skills and confidence (van Haeften-Van Dijk, van Weert and Droës, 2015; Windle *et al.*, 2019).

Art is not a panacea, however. The production and consumption of arts and culture can enforce power and reproduce inequalities, leading to marginalisation of those who do not possess the right sort of cultural capital (Bourdieu, 1984; Belfiore and Bennett, 2007).

Nevertheless, at its best and for those able and inclined to access it, indications are that art can, as Tolstoy suggested, be a ‘means of union’ between people. It can provide a way of ‘joining them together in the same feelings’ – making connections, reinforcing identities, and providing social support when and where it is most needed (Dupuis *et al.*, 2016; Skinner *et al.*, 2018). Even if this is all it is, Tolstoy might perhaps have argued, this might be sufficient for arts activities to claim a place in the toolboxes of those seeking to understand, support, and care for people living with dementia.

However, when it comes to commissioning of health and social care services, there is little space for exceptionalism. Those who champion the arts must also make a case for their effectiveness. For dementia care and treatment, this means competing for limited resources alongside other interventions, both pharmacological and psychosocial. In this competition, the evidence derived from evaluation and research is considered crucial in helping those holding the purse-strings to invest their resources wisely. Providing this evidence has – however – not been found to be easy.

1.2 Justification for the research

The subject of this thesis is the methodological challenges that face evaluators of arts-based activities for people living with dementia. The research aims to identify and describe the challenges as they are reported within the literature, and to explore how they are experienced in the practice of evaluation stakeholders – unpicking the problems. Understandings gained will be used to signpost potential solutions and to recommend directions for future practice and research.

Research and practice related to the arts and dementia situates itself within a wider field termed arts and health. Here, arts and cultural activities have been claimed to be effective in addressing a diverse range of health and wellbeing needs for the population as a whole. Allied to arts and health are the various arts therapies. There is also a current of thinking that regards access to the arts and creativity as beneficial to us all, regardless of age or of diagnosed health condition (Cutler and Bakewell, 2009).

There have been consistent demands for greater research rigour for arts and health and its allied fields, including through application of robust controlled methods that would address a perceived need to evidence precise and measurable health outcomes (Skingley, Bungay and Clift, 2011; Clift, 2012). These methods have proved difficult to apply within the open and variable settings in which the arts are practised. It has been argued that theoretical and conceptual frameworks are urgently required to support such work and allow a greater understanding of its processes and underlying components (Stickley *et al.*, 2017). A growing pressure towards engagement with the arts for benefits, including health and wellbeing, may be being encouraged without sufficient formulation around identified need, theories of change, anticipated outcomes or alternative approaches; with detrimental effects for practice and for those who might stand to benefit from arts engagement (Phillips, 2019). Although there are indications that research attention may be turning in these directions (Fancourt and Finn, 2019), much evaluation still focuses on the assessment of outcomes and impact.

There have also been concerns that claims for what the arts can achieve may be exaggerated because of pressure towards application of experimental research paradigms to questions about their effectiveness (DeNora and Ansdell, 2014), or out of a desire to advocate to the benefit of arts organisations delivering the work (Goulding, 2014).

Significant research attention has been addressed towards the arts therapies, whose structures and goals are comparatively well established (Beard, 2011; Cowl and Gaugler, 2014), although they have proven similarly difficult to evidence convincingly (Vink *et al.*, 2011; Karkou and Meekums, 2017; van Der Steen *et al.*, 2018).

Within arts and health, arts and dementia has been termed an emerging field (Zeilig, 2016). Historically, as in arts and health generally, the majority of research and evaluation has comprised small scale studies examining the efficacy of short-term interventions. However, recent developments in the UK have sought to establish stronger foundations (Camic, Zeilig and Crutch, 2018). Examples have included a project seeking to understand the theoretical underpinnings for visual arts programmes (Windle *et al.*, 2018), research identifying the principles and features that might inform a 'taxonomy' of arts interventions (Cousins, 2018; Cousins *et al.*, 2019a), and a programme contributing research and activity intended to enhance public understanding of dementia and the creative potential of those who are living with the condition (Camic *et al.*, 2018; Zeilig *et al.*, 2018). Positively, a recent scoping review conducted for the World Health Organization has described the evidence base supporting the implementation of arts for patients with dementia as 'significant' (Fancourt and Finn, 2019).

Research methods routinely used and validated for other populations can pose difficulties when applied with people living with dementia. Interviews, for example, rely on participants being able to respond to discussion of abstract concepts, to use recall skills, and to report opinions verbally – all things that can be difficult for those with the condition (Beuscher and Grando, 2009). Cognitive frailty, issues with consent and capacity, and stigma contribute to these difficulties and have led to people with dementia being routinely excluded from research about them, or to their voices being represented by proxies (Hellstrom *et al.*, 2007).

Some of the most commonly-claimed outcomes of arts practice, such as enhanced subjective wellbeing or quality of life, are widely contested and poorly understood in the context of dementia. They are difficult to measure in the face of diminished cognitive capacity and memory (Bowling *et al.*, 2015).

The measures and methods chosen by a researcher or evaluator will reflect the underlying conceptual frameworks to their work. In the case of arts activities for people with dementia, they can sometimes reveal a tendency to privilege a bio-medical understanding of the condition. A recent review of the indicators used to evaluate the impact of music on the health and wellbeing of people with dementia revealed that the reduction of neuropsychiatric symptoms has received the greatest attention from researchers in this context (Dowson, McDermott and Schneider, 2019). While such a focus is understandable, there is a recursiveness to it, making it likely that less attention will be paid to other potential benefits that the arts might also provide. It is also possible that application of the methods of medical or pharmaceutical science may be over-simplifying the tensions and the messy ontological, epistemological, and methodological issues that litter any field attempting to bridge the humanities and the sciences. These issues include difficulties in understanding what actually happens when we *do* art, what any given individual might *value* about the activity, or indeed what it means for an individual to *feel healthy*, to *live well*, or to experience the opposites of these.

As researchers begin to appreciate the imperative upon them to ensure that the voices of people with dementia are heard in wider society (All-Party Parliamentary Group on Dementia, 2019), the continuing existence of distinct barriers to their involvement are increasingly recognised as contributing to methodological failures in research. From the researcher's point of view such barriers include the resources and time required to enable participation, ethical concerns and negotiations with existing ethical processes, as well as difficulties in identifying and recruiting participants. People with dementia sometimes perceive the research process as complex or feel they lack the expertise to contribute, and there is always the potential for research to cause distress (Bethell *et al.*, 2018).

While the ultimate goal of research is to produce new knowledge, evaluation tends to more practical and pragmatic ends, supporting the needs of policymakers and programme managers through investigation of the value, effectiveness or mechanisms involved in a particular programme or activity and its implementation (Clarke and Dawson, 1999). Evaluation has political effects and will be influenced by political forces. Evaluators are routinely required to negotiate issues including those relating to the use of their work, whose questions are to be addressed, and whose interests served by an evaluation and its findings (Weiss, 1979).

In the field of arts and dementia, evaluation effort has sometimes been directed towards the need to build an evidence base. However, most providers of arts activity for people with dementia are small, often third sector organisations who struggle to evaluate their activities according to the principles of evidence-based practice because they lack financial resources, technical skills, and evaluation literacy (Daykin *et al.*, 2017; Bach-Mortensen and Montgomery, 2018). Their evaluations are much more likely to be produced as part of an effort to demonstrate their accountability to funders. Like researchers, evaluators working in this way also experience friction when faced with selecting methods to fit both the health context and that of the arts (Fancourt and Joss, 2015).

Because of the context in which evaluation is produced and used, its methodological challenges – although they are not separate from those experienced by researchers – may require different framing. The promise of research and practice in the field of arts and dementia is that it will create more, or more *effective* provision of access to quality arts activities and help to realise their potential benefits for health and wellbeing. If these promises are to be kept, the challenges for their evaluation, including how these compare to those facing researchers, must be better understood and addressed.

1.2.1 Arts, health, cultural policy and the politics of evidence

Despite historical attempts to separate the methods used to access scientific truths from the values that might determine the focus of research, it is increasingly argued that all science is essentially and inescapably value-laden (Kelly *et al.*, 2015; Elliott, 2017). We engage in evaluation and research and employ methodologies to access knowledge for many reasons, some of which have to do with how we reflect and interpret the different values inherent within the cultures and structures that surround us. A search for evidence and the judgements of value that accompany it cannot be disinterested.

As has been outlined above, the epistemological and associated methodological controversies and tensions of the field of arts and dementia reflect those within its parent field of arts and health, as do associated calls for greater methodological rigour. They also relate to inquiry into the value of the arts and culture for society as a whole (Crossick and Kaszynska, 2016). Here there has been significant pressure for methods of evaluating arts impact to move towards production of evidence that would render arts and cultural activity more easily auditable. This has been particularly the case where the aim is for it to contribute to public health improvements, or play a role in health and social care practice (O'Brien, 2010). This thesis is itself part of a research programme, funded by the Alzheimer's Society, whose aims include a desire to strengthen the evidence base for use of the arts in the care and treatment of people with dementia. Again, the politics of evidence sit at the heart of all such work: specifically, the pressing requirement for research and evaluation of arts practice to produce results that can usefully inform or influence policy and associated investment (Clift, 2012).

However, so far, effort put into such activities has not been wholly successful. A 2017 report commissioned by the UK All Party Parliamentary Group on Arts, Health and Wellbeing presented multiple examples of the effectiveness of the arts in improving health and wellbeing. But in the report's foreword, Alan Howarth wrote: '[T]he conundrum that we have found ourselves pondering is why, if there is so much evidence of the efficacy of the arts in health and social care, it is so little appreciated and acted upon' (All Party Parliamentary Group on Arts Health and Wellbeing, 2017, p. 5).

In order to participate within the discourse that speaks to and with policy, assessment of the *arts* side of the arts and dementia equation has tended to be focused towards its capacity to produce particular impacts in a cost-effective manner and thus (it is generally supposed) to address the needs of policymakers. In relative terms, less attention is paid to its less easily auditable intrinsic qualities – those that relate to its subjective, individually, or culturally-produced values. In the Creative Health report cited above, Howarth rejects argument between these two sides as a 'chronic and sterile altercation', while claiming rather mysteriously that 'the validity of art itself' can lead to better health and wellbeing. However, this refusal to engage with or further diagnose the symptoms of a perennial argument could be seen as part of the problem.

The Creative Health report was created out of a desire to demonstrate the existing evidence for arts and health and to outline a persuasive foundation for future evidence-building; these are its stated aims. In rejecting a need for further discussion about the value of art and the nature of the evidence that is required to validate its contemporary existence, the author of its introduction also fails to question what it means for the field of arts and health practice, and evaluation research about it, to be shaped by the epistemological framework supporting evidence-based practice.

The term *evidence base* reflects a particular set of ideas informing research study designed to have impact in the public sphere. It has, over the past half century, been applied to an increasing number of fields of specialised activity, although it originated in medicine. Evidence-based medicine (EBM) sprang out of the concern of epidemiologist Archie Cochrane in the 1970s with what he saw as potentially dangerous practice around the use of clinical procedures and processes. It has been defined as ‘the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients’ (Sackett et al., 1996, p. 71). Cochrane called for the systematic, unbiased evaluation of all clinical treatments, ideally through the use of randomised controlled trials (RCTs) or synthesis of results from these through meta-analysis and systematic review. The result has been to cement into general understanding a hierarchy of different kinds of evidence, with meta-analysis and review, along with RCTs, at the top, and case studies and practitioner knowledge at the bottom.

The EBM movement has positively transformed a multitude of areas of clinical epidemiology and healthcare generally. It has led to the creation of guidelines and frameworks for medical and care practice by many bodies, including the UK’s National Institute for Health and Care Excellence (NICE). And it provides the theoretical underpinning to support what is the standard current approach to evaluation of all health-related interventions, exemplified in guidance in the UK produced by the Medical Research Council (MRC) (Craig *et al.*, 2008).

In recent years, as far as culture and the arts is concerned, and despite nuanced discussion among epidemiologists, clinicians, and public health specialists, evidence-based practice remains a significant goal for any arts activity that might seek to create social value. The arts sector is consistently encouraged to respond to twin requirements: provision of ‘the hard evidence that a largely scientific community, like health, demands’ and evidence of value for money for commissioners (Slay and Ellis-Petersen, 2016).

Such encouragement has been influential in determining the shape of evaluation frameworks used in the arts and health sector, particularly those designed to assess wellbeing impacts. In 2016, Public Health England (PHE) issued a framework designed to support evaluation of arts, health, and wellbeing projects in response to an identified need to provide ‘robust evidence’ for the ‘effectiveness, impacts and costs’ of the arts to encourage their inclusion within the commissioning of health and social care services (Fancourt and Joss, 2015; Daykin and Joss, 2016).

The major funding bodies, including the regional Arts Councils, also commonly provide their own evaluation guidance.

An appreciation for the principles of evidence-based practice in evaluation, alongside awareness of contemporary debate around its limitations (Goldenberg 2006; Ashcroft 2004), has informed my research. The debate demonstrates that, for evaluators, difficulties may not be confined to practical challenges relating to the more rigorous application of method, or even to tensions resulting from disciplinary or paradigm differences. The field also faces fundamental difficulties in determining the purpose and in enhancing the utility of evaluation practice and its results. As a result, the task of identifying solutions is unlikely to be as simple as plumping for one epistemological outlook or another.

1.3 Aims and objectives of the research

This study aims to contribute towards effective evaluation practice in the field of arts and dementia. Its ultimate beneficiaries will be people living with dementia, but in reaching this goal, it is hoped that its findings will support the work of evaluators of all kinds, as well as that of artists and creative practitioners and arts organisations.

In order to achieve this aim, I will identify and conceptualise the methodological challenges reported by evaluators and researchers in the field through a literature review. I will explore these challenges and situate them in the context of the wider stakeholder experience of evaluation through a series of interviews and analysis and discussion of the resulting data. I will then use the understandings reached to recommend solutions for evaluation practice.

1.4 Defining terms

For a researcher, a clear definition of terms can be helpful in providing useful boundaries and structure to a project. It can also help readers to understand a research project’s scope and aims. Definition becomes particularly important when an enquiry concerns terms about which it may be assumed there is common understanding. In this project, such terms include *methodological challenge*, *arts-based activities* and *evaluation*.

1.4.1 Methodological challenge

Methodology has been defined broadly as ‘a way of thinking about and studying social reality’ (Strauss and Corbin, 1998, p. 3), and perhaps more conventionally as ‘the strategy, plan of action, process or design lying behind the choice and use of particular methods and linking the choice and use of methods to the desired outcomes’ (Crotty, 1998, p. 3).

While the etymological origin of the suffix *-ology* should point us to an understanding of it as *the study* of methods, it is often conflated with their (correct) application. For my research I have chosen to highlight Abraham Kaplan’s definition of methodology as ‘the description, the explanation and the justification of methods’ (1964, p. 18). This definition draws attention to the work involved in methodology, and to what I see as its subjective, contextual and dialogical nature. To Kaplan, I would add the further interactive dimension of implementation or application of method.

In the construction that I use, methodological work is not a series of closed or one-dimensional actions. Instead it is a process that involves and necessitates critical interaction with the world in which methods are applied, and with the structures shaping that world. The challenges faced by researchers in engaging with methodology are never just theoretical or just practical. Methodological decisions, along with their description, explanation and justification, are made by individuals viewing the world from particular perspectives and acting in ways that reflect these perspectives, present contexts, and past experiences. Further, it has been suggested that research methods are social and essentially performative, that ‘they have effects; they make differences; they enact realities; and they can help to bring into being what they also discover’ (Law and Urry, 2004, p. 393).

This research project is informed by an understanding of methodology as occurring within, and being affected by, particular social, political and historical contexts, of being in dialogue with past and present iterations of itself and other methodologies, and of being shaped by the subjectivity of the researcher. With

Law and Urry, I see methods as capable of having real effects within the real worlds in which they are applied. In my understanding, it is from these effects and this interaction with the realities of the world that methodological challenges for evaluation will result.

Some of those effects might be detrimental to the conduct of an evaluation itself, placing barriers in the way of achieving accurate or effective results. Some could even be detrimental to the work of the programme under evaluation, or those who stand to benefit from it. The barriers facing evaluators may therefore be both theoretical and practical. They may have their origin in any stage of an evaluation or research process, from its inception to its dissemination and translation into practice. Some are likely to be clearly reported by authors in their consideration of an evaluation study's limitations. However, some authors, particularly those working within interpretivist or anti-positivist paradigms, may not report limitations directly. Outside the academic world, arts and dementia evaluators may not report them at all, creating difficulties for those reading or acting on their work to know whether they can have confidence in its results. Furthermore, we may not be able to understand whether there are difficulties that are being felt most keenly by those who are not responsible for authoring evaluation studies – in the case of this study, those who commission or participate in the projects being evaluated. One of the hypotheses underpinning this research was that there might be barriers to the application of methodology that are going unreported and even unrecognised.

1.4.2 Arts-based activities

The focus of this study is arts-based activity. This is defined as: an active engagement with the expressive components of visual art, music, drama, dance or literature, with or without the facilitation of a professional artist, for leisure purposes, enjoyment or the encouragement of social interaction and other health and wellbeing aims. This definition could be used to describe much of the practice currently being delivered in the field of arts and dementia.

For my purposes it does not include the work of arts therapists. Most commonly *therapy* is thought of as activity delivered by a trained practitioner, often –

although not always – in a clinical setting, or with a clinical goal in mind (Beard, 2011; McDermott *et al.*, 2013). However, there are confusions in the field about what might be termed *therapeutic* activity and what constitutes therapy itself (Cousins, 2018). A differentiation between studies that have 'wellness' as their focus and those that are focused on 'treatment' was made by the authors of one review of participatory arts for older people, although – arguably – this served only to support a potentially damaging assumption that 'wellness' is not a goal to be applied to people living with dementia (Noice, Noice and Kramer, 2013). Even if an arts activity is not specifically labelled as therapy, research into the arts for people with dementia has often focused on the potential for creative activities to treat or manage therapeutically adverse symptoms of the condition (Bungay *et al.*, 2014; van Der Steen *et al.*, 2018) and it is not always clear why this potential might not also be considered therapy.

However, the range of practice being delivered in the field suggests that people living with dementia engage in the arts in order to address a variety of needs, not all of which will be clinically expressed. In this context, the absence of a clear distinction between arts therapy and arts practice has been identified as placing an 'unreasonable burden' on arts practice (White, 2009; Broderick, 2011). It may also cause particular difficulties in researching the subject when an arts activity involves those who are living with a defined medical condition, but whose main concerns are to live better, or to receive care that might enable this while supporting their identity and their status in relation to others in society – their *personhood* in other words (Kitwood, 1997; Brooker and Latham, 2016).

For the purposes of this research, my decision was that the goals and processes of arts-based activities are likely to differ from those of therapy, as will the conditions and settings in which they are conducted, and – importantly for evaluation – the structures through which they are commissioned. I felt it useful to recognise the training and experience of therapists as different to that of professional artists or other facilitators of creative activity, such as trained dementia care staff. These elements are of importance to consideration of how any given activity will be implemented and evaluated. Therefore, in this research I made an early and conscious decision to consider as out of scope the evaluation of practice specifically termed therapy. It should be noted however, that, regardless of these differences, many of the challenges for evaluation of arts therapy for people with dementia will be the same or have similar origins. Therefore, the recommendations and implications of this study should have resonance for this field as well.

The term *intervention* is commonly used in the research literature to describe a planned and purposeful action by an individual, group or other entity, designed to create change (Midgley, 2003). Associated with application in the context of experimental evaluation and research study, its use is aligned with paradigms familiar from the sciences. It has been termed helpful in indicating that a given kind of art practice has underlying purpose and theory, that it is not simply 'doing art' in some random way (Cousins, 2018). However, its use is contested (Parkinson, Windle and Taylor, 2017), and the term is not generally applied outside the academic context, except where evaluators or artists are seeking to emulate academic terminology, or perhaps to help convey academic integrity. I felt it important in this research to use a term that could more accurately reflect the contributions made by all of those involved in the practice of arts involving people with dementia, including the interactions of those who are living with the condition.

Use of the word *activity* serves to highlight a difference from engagement that is passive or non-participatory. It is also useful to recognise that people with dementia, like those without the condition, will engage in the arts purely for personal enjoyment, rather than to address defined psychosocial needs. For these reasons I felt that *activity* provided a more neutral and democratic way of describing the kinds of arts practice explored in my research.

While it is beyond the scope of this study to explore or define what is and is not *art*, it is also worth noting why the term *arts-based* has been used to describe the kinds of activities being evaluated. This is intended to reflect the way in which many activities in which people with dementia participate will have the arts as their inspiration or catalysing force, but do not confine themselves in their application to the practice of any artform in particular. For example, in many museum and gallery-based cultural programmes currently delivered, the arts provide a foundation for participatory activity that is meaningful to those involved, but whose content ranges widely, incorporating educative elements, sensory stimulation, or reminiscence and life-story work (Camic, Tischler and Pearman, 2014; Nyman and Szymczynska, 2016; Thomson *et al.*, 2018).

For the purposes of my study, I am concerned only with activity that makes use of the capacities of the arts for expression. Activities that do not do this might include those where music is used purely as the backdrop or stimulus to exercise or during personal care, social outings to the theatre or concerts, or the application of visual arts principles in the design of healthcare facilities. I have also excluded activities such as colouring-in exercises designed to calm anxiety, some aspects of cognitive stimulation therapy, and some forms of craft. While many of these will have value, they do not necessarily involve individuals in *making art* in the sense used by philosopher and psychologist Dewey when he described aesthetic experience as a potentially transformative 'act of expression' that also involves appreciation, perception and enjoyment (Dewey, 1934).

It has been suggested that we need to 'broaden the debate' on creativity and arts (Bellass *et al.*, 2019) and there is a growing and important interest in everyday creativity (Richards, 2007; Silvia *et al.*, 2014). This includes recent well-justified calls for exploration of whether arguments for the value of the arts reflect a proscriptive 'selective tradition' in which particular kinds of cultural or leisure activity have been privileged above others (Oman, 2019). There is much to recommend this argument, but neither 'everyday creativity' nor 'everyday participation' were the subject of the current study.

Because *arts-based activities for people living with dementia* is too lengthy a term to be reading constantly in an extended piece of writing, where it sometimes might be clumsy or long-winded to use it, I have replaced it instead with the portmanteau phrase *arts and dementia*. For the current purpose, please take them to mean the same thing.

1.4.3 Evaluation

As was mentioned earlier, evaluation has a different focus to research, perhaps falling within the remit of 'applied research'. Its primary purpose has been described as 'not to prove but to improve' (Clarke and Dawson, 1999), although this does not always reflect its use in practice. Lincoln & Guba have described the act of evaluation as that of constructing an interpretation or ‘making sense’ of a situation (1989). Its purposes are usually achieved through some form of systematic data-based enquiry into a particular subject, programme, or activity. Evaluation methods and goals are likely to differ according to whether the aim is to establish and assess the value of a particular programme of activity, or to discover how a programme works with the aim of supporting its future implementation in similar or different contexts. Evaluation can be summative (designed to assess how effective a particular delivery has been) or formative (designed to shape delivery) (Scriven, 1967). It may be focused on either process or outcomes, and sometimes both of these in combination (Chen, 1996). As with research, its aims and purposes will guide its questions and therefore its methods, and may also determine the nature of the role assumed by the evaluator and his or her relationship to the body or individuals commissioning the evaluation (Guba and Lincoln, 1989).

However, even those who call themselves evaluators do not necessarily define evaluation in the same way, and researchers may define it differently still (Wanzer, 2019). Rather than assuming that evaluation means one thing or another to those who are involved in evaluating arts-based activities for people with dementia, this study sought to find out what the experiences of those people tell us about evaluation in this context. It may be that evaluation is misunderstood or understood in very particular ways by stakeholders. Reaching a more accurate understanding of the term for the field in the light of this experience could be seen as – in itself – one of the aims of this research.

1.5 Stakeholders and impact

The stakeholders for this research are also those for evaluation of arts and dementia activity. It aims to provide information to support evaluators and researchers conducting future evaluation of arts and dementia activity. It recognises the valuable role played by artists and those involved in managing artists in delivering and supporting evaluation activity, alongside that of health, dementia care, and other professionals who work with, or provide an enabling context for arts practice. Its ultimate beneficiaries will be people living with dementia and their families and loved ones who will reap the rewards resulting from provision of more effective, better understood, and more accurately targeted and meaningful arts practice.

1.6 Thesis structure

In this Introductory chapter I have provided context and background to the research, making a case for why it is necessary, and defining some key terms.

Chapter 2 presents findings from a narrative synthesis of the literature seeking to identify the limitations and challenges for research and evaluation in the field of arts and dementia, as reported by authors of published evaluation and research studies. A number of themes underlying and connecting these findings are identified. These formed a set of concepts that were then used to sensitise data collection in the next stage.

My methodological approach and theoretical influences are outlined in Chapter 3. I present the process by which these were selected and discuss the relationship between them. The approach to analysis and theory development in this research was derived from and inspired by a blend of critical realism and grounded theory. The fundamentals of both elements are described, and a case made for their suitability for addressing the subject of this research.

In Chapter 4 I describe the way in which my chosen methodological approach was used to inform conduct and analysis of a two-phased series of stakeholder interviews. Implications relating to ethics, reliability, and reflexivity are identified. In the first phase I used grounded theory methods to guide selection from a pool of potential participants initially identified through a grey literature search. They included artists, arts organisation representatives, funders and commissioners and evaluators and researchers. I applied a continuous and iterative analytical process and conducted second phase interviews with experienced evaluators and researchers in testing and refining the analytical categories derived. Towards the end of the analysis process, I worked with an artist in creating a set of narrative graphics intended for use in disseminating the research. The resulting collaborative approach to creative dissemination and its contribution to the analysis process are also described.

The findings from my analysis of a set of 26 stakeholder interviews are presented in the fifth chapter. I identified thematic categories, and these are described in detail. The graphic narrative illustrating my findings for an audience of arts practitioners and the general public is also included in this section. My findings are further discussed in Chapter 6, synthesised with the findings of the literature review, and situated in the research and policy context to suggest a response to my primary research question.

Chapter 7 summarises the study's findings and uses these and the associated discussion in signposting solutions to the challenges that have been identified. I reflect on the research process and outline implications and recommendations for future research.

In Chapter 1 I established the background to this study and introduced the aims and the justification for the research. In the following chapter I outline my approach to reviewing the literature. I present the findings of a review seeking to identify the limitations and difficulties for research and evaluation as reported by authors of published studies, and to understand what this might tell us about the methodological challenges evaluators and researchers face. This review also helped to inform my research in ways that I will go on to describe in Chapters 3 and 4.

Illustrations used throughout this thesis were created by Emma Lazenby of ForMed Films CIC as part of a collaborative project to disseminate the research.

2 LITERATURE REVIEW

Conducting this literature review helped me towards an understanding of the methodological tensions and difficulties for evaluators and researchers of arts and dementia activities. It allowed me to identify what they report as the limitations and challenges of working in the field. Through an iterative and interpretive synthesis of the literature, and using a hermeneutic framework, I developed a provisional set of *sensitising* concepts. These are woven through the remainder of the thesis and support data collection in the following phase of the research.

2.1 Review aims

In this thesis, I have followed Abraham Kaplan in defining 'methodology' as 'the study – the description, the explanation, and the justification – of methods' (Kaplan, 1964, p. 18). To this definition, I would also add the implementation or application of methods (see section 1.4.1 above). In reviewing the literature, I wanted to know what evaluation researchers say are the challenges for successful evaluation and research practice. I was also interested in developing an understanding of how these might be linked to the conditions in which evaluation is conducted. I hypothesised that there might be methodological hot or blind spots, and areas where challenges might be unreported or unacknowledged.

2.2 Review methods

There are many kinds of literature review (Grant and Booth, 2009). At what is considered to be their most rigorous, under the rules of the Cochrane Collaboration for example, systematic reviews and meta-syntheses are conducted with detailed protocols, pre-specified questions, explicit and replicable search strategies, tools to assess study quality, and defined methods that combine, summarise, and otherwise analyse or synthesise findings from included studies (Egger, Smith and O'Rourke, 2001). Transparency allows replicability of review processes in a way that is considered crucial when the aim is to inform evidence-based practice. However, alternative kinds of systematic review are sometimes acknowledged as more appropriate (Greenhalgh, Thorne and Malterud, 2018), particularly where there is a need to include qualitative research, studies reflecting different research paradigms (Booth, 2001), or if the aim is to explore concepts and the connections between them (Suri, 2013). Rigorously conducted systematic reviews often benefit from at least two authors supporting and checking their assessment and extraction of data and reporting of results (Boland, Cherry and Dickson, 2017) – a resource unavailable to a lone PhD student.

There is a small body of literature relating to methodology for researching the arts and dementia, however a review that only included this would have been limited. This is because the discussion of methodological challenge for evaluators is largely integrated within the literature reporting on evaluation and evaluation research studies itself. Although the field of arts and dementia is relatively young, still any systematic search process encompassing it as a whole will reveal a large and unwieldy body of literature, including material reflecting a variety of research paradigms and of varying standards of quality.

Additionally, not all evaluation conducted by academic teams results in publication in peer-reviewed journals. Much evaluation is conducted outside the academic world, by independent evaluators and consultancies, arts and third sector organisations, or by artists themselves (Daykin *et al.*, 2017). Here it is acknowledged to be difficult to conduct exhaustive reviews; both qualitative and grey literature have been described as 'elusive' (Dixon-Woods, 2006).

This review was not intended to produce generalisable findings for application in evidence-based practice. Instead, my aim was to engage intellectually with the literature and to classify and then interpret its contents in a way that would help develop my own understanding of the problems in the field as well as that of readers of this thesis. This literature review played an iterative part in the progress and processes of the thesis, rather than being just one step in a pre-defined or linear journey. Taking these considerations into account, the overarching framework I chose to apply was a hermeneutic one (Boell and Cecez-Kecmanovic, 2014). This approach emphasises continuous engagement with a body of literature, rather than a strict reliance on replicable systematic method or comprehensive searching.

Nevertheless, there were systematic elements to the process. I conducted a review of studies published in peer reviewed journals, using systematically-informed search methods. This was followed by a purposive search and rapid review of grey literature to help me check for insights I might have missed as a result of looking only at published studies. I also used this search to identify interview participants (see section 4.2.2). After data collection and analysis were complete for the study itself, I updated the review with additional searches for published literature.

Given the nature of the question driving my research and the hermeneutic framework I was planning to apply to the process, I determined a narrative and interpretive approach appropriate to the synthesis. I accepted that the review was not going to be exhaustive, and that my aim was exploratory, seeking connections between material. I have written and re-written this chapter a number of times, feeling that each re-writing is helping me get closer to an understanding. In defining the purpose and use of critical interpretive synthesis review methods, Dixon-Woods identified clarity of process, an open and reflexive approach and the careful examination of decisions and statements as key to producing good results (Dixon-Woods, 2006; Dixon-Woods et al., 2006). My aim was to apply these standards to my review.

2.3 Review question

Following Dixon-Woods, I made the decision to allow a review question to develop iteratively. The question that guided my initial search and data extraction was a ‘compass rather than an anchor’ (Eakin and Mykhalovskiy, 2003). It was:

What do authors of studies evaluating participation in arts-based activities for people with dementia report as the challenges for, and limitations of, their evaluation research practice?

However, the question that went on to shape the final synthesis and reporting of this review was:

What can we understand from a review of evaluation and research studies about the challenges facing evaluators of arts-based activities for people with dementia?

I describe the process by which I moved between these questions, and why, below (section 2.3.3).

2.3.1 Inclusion and exclusion criteria

I did not narrow searches down by artform, setting, or by type or stage of dementia. I used a wide definition of *evaluation research* to identify published papers and to include research studies reporting on relevant arts activity. Arts and dementia is a field in which multiple disciplines meet, and therefore the research and evaluation connected with it reflects a variety of paradigms, epistemologies, and theoretical approaches. I was interested in the challenges apparent from exploration of the field as a whole and so I have included studies representing this multiplicity. The use of a quality assessment tool would have been problematic given the varied methodologies of included studies. More importantly, I did not consider it appropriate to apply quality standards. Because I was interested in the process of evaluation and research rather than its results, even those studies whose methodological quality might be considered ‘low’ were likely to be relevant.

Nevertheless, inclusion and exclusion criteria could and did structure the search. I included only literature related to arts-based activities practiced with or for people living with dementia and which was either an original research study published in full, or whose discursive content directly related to the methodologies used for evaluation of a relevant activity. I excluded studies dealing solely with the impact of arts activities on family carers, professional caregivers or care organisations, artists, arts and cultural organisations, and members of the public. I also ruled out studies relating to the arts as practiced with older people more generally, unless it was clear that the majority were living with dementia. I excluded experimental studies involving an arts element used solely to test clinical symptoms, such as cognition or recall, and studies exploring the symptoms and experience of artists themselves living with dementia.

In line with the definition provided earlier (section 1.4.2), I distinguish between *activity* and *therapy*: this research focuses on activity. Because the factors that distinguish the two may impact both on method and evaluation aims, studies involving the various arts therapies were specifically excluded from this review.

I differentiated between *active* and *passive*, excluding any literature in which participants are passively in receipt of the arts or culture, for example: caregiver singing during personal care situations to decrease resistance to care; the playing of recorded background or soothing music during care activities; personalised playlists in which a participant plays no active role in selecting the music, or; literature that is about the deployment of art and design elements within a care environment. I also excluded studies involving live performance by dancers, actors or musicians which might be described as *entertainment* and where the performance was not intended to lead to any active engagement beyond simple appreciation.

Because this is a young field, I set a date limit of studies published in and after 1998 and took searches right up to the date of the search (February 2016 in the first instance and October 2018 for the updated review). For practical reasons, I excluded literature not in English.

2.3.2 Search strategy

Initial searches were first conducted with the support of the subject librarian at the library of the University of Worcester in February 2016 using three electronic databases: CINAHL, PsycInfo and Medline. These were chosen because they offered a broad range of titles likely to be of relevance. Titles of journal papers were searched for the following search categories, adapted for use in individual databases: (dementia OR Alzheimer's) AND (art OR arts OR music OR danc* OR drama OR singing OR poet* OR theatr*). I extracted and scanned titles and abstracts, removing duplicates and items that did not fit the inclusion criteria. I downloaded 84 full text studies. Following further exclusions (e.g. not subject of interest, not an evaluation research study, not in English, not an 'activity'), I identified a total of 60 studies for detailed reading and analysis, which began in June 2016. Fourteen relevant reviews were also identified and provided background for analysis but were not included in reporting.

I updated the review in October 2018, performing the same searches but limiting the dates to the range March 2016 – October 2018. At this point I also performed a simple Google Scholar search for 'arts AND dementia AND evaluation', going through the first 20 pages of results. From both these sources, and following the processes described above, I identified a further 24 relevant papers for data extraction. Because the number of additional papers was extensive, this new material was incorporated into the synthesis through an update. Rather than substantially altering the analysis from the first review, it served to confirm and add new examples.

2.3.3 Data extraction and synthesis

I read and extracted data in stages. First, I read each paper, paying detailed attention to the sections where limitations and methodological issues were reported by the authors or where they described challenges or made recommendations for future research and evaluation. I extracted all examples of these in note form into a bespoke tabulated Word document. In this I also noted details of the study design and methods, chosen methodology, the aims and objectives of the study and the reported limitations and methodological challenges. This document formed the basis for my table of characteristics. Sometimes the extracted information was not signposted in a limitations section, particularly if the approach was qualitative. If this was the case, I read the paper in greater depth, but still took care to only include issues that authors reported explicitly to be problematic.

Regardless of research paradigm, most (although not all) papers included methodological limitations or some kind of critical or reflective discussion of the methods used. Some limitations sections were formulaic, while other authors included thoughtful detail in their discussion. In reporting limitations, it was notable that authors tended to reference concerns fitting a broadly positivist model for the conduct of evaluation research, even when the study itself was sympathetic to or used a different approach.

To analyse these data, I first used a general inductive approach (Thomas, 2006), guided by my initial question. Using the data extracted previously, I created a table of characteristics including an entry for each study, in which I noted the study aims, methods and methodology, stated limitations or methodological challenges, and further notes about the study of relevance to the research question (authors' methodological recommendations for future research, for example). To produce a thematic synthesis, I re-read this table of characteristics, identifying and noting common themes as I did so. I then categorised these into higher level thematic areas and wrote descriptions for each (see Tables 2, 3 and 4 below).

This process gave me some clear answers to my initial question. On their own, however, these told just part of the narrative since they were only the reported issues and were framed by the concept of what is generally considered a 'limitation'. This reminded me that the concept of methodological challenge and what constitutes a limitation or barrier to successful evaluation practice is likely to be a reflection and construction of an evaluator's own epistemological perspective. Reporting bias, publication bias, or dissemination bias could also be factors: authors may be unable or reluctant to report on studies considered flawed, or where the results are unclear, and academic journals have strict criteria which they apply to submitted work before it is accepted for publication (DeVito and Goldacre, 2019). This is likely to have led to homogeneity in what was being described as a limitation or a challenge.

It was also possible that authors might be reporting limitations and challenges not because they felt them to be genuinely limiting factors, but because of transferred or assumed quality standards (Eakin and Mykhalovskiy, 2003), some of which might be inappropriate in the emerging multidisciplinary field of arts and dementia. In addition, much evaluation of arts and dementia practice takes place outside the academic context, sometimes as academic consultancy rather than funded research (Daykin, Attwood and Willis, 2013). It could be that evaluators working in this way would report different challenges, or conceptualise challenge differently, just as they do evaluation itself (Wanzer, 2019).

The review's initial findings had therefore identified what authors were reporting as limiting or problematic, but went no further in illuminating the causes of these limitations. Deeper exploration of the content and context of the material under review was required, and this meant a switching of analytic mode from 'aggregation' to 'interpretation'. That is, it entailed moving from the summarising of secure and well-specified concepts to the development of concepts and theories that might integrate those summarised concepts (Dixon-Woods et al., 2006).

To build on the initial analysis, and to make it useful in guiding the remainder of my research study, I returned again to the literature. I re-read the original papers with my categories in mind, paying close attention to where there might be links between categories and themes that were latent rather than semantic. For each study, I recorded further notes as to how the already extracted and synthesised data might relate to the authors' intended aims for their work and their recommendations for future evaluation practice or areas of study. I applied a second level of coding for conceptual themes and these were discussed with my supervisory team and other colleagues. Initial reflections were presented at a conference and published in a journal for comment and discussion (Gray *et al.*, 2018).

The result was a set of five 'sensitising concepts'. These are defined as concepts that might help in the consideration of, or questioning of what it will be useful to consider about, my research question (Charmaz, 2014). In Chapter 4 I describe how these informed my primary data collection.

2.3.4 Grey literature

After finishing analysis of the published papers, I performed a rapid grey literature search in February 2017. I define ‘grey literature’ as material in print or electronic format that is not controlled or published by commercial publishers (Godin *et al.*, 2015). The search was limited to evaluation reports or studies of arts and dementia activity programmes, written in English that had been disseminated by a funder, commissioner, evaluator, or provider of arts-based activities for people with dementia in the UK. Doctoral level dissertations passed by an awarding body within the UK were also included. The search had four strands: (1) a search of the Dissertation Abstracts database; (2) a Google search; (3) a search of the downloadable contents of a set of targeted websites; (4) a cascade search informed by expert consultation and interview. I set date limits of 2006 – 2016 because web searches would be unlikely to recover material of an earlier date.

Applying the same criteria as for the published literature, I included 29 items, and extracted data using the same templates and processes. I created a similar table of characteristics and coded for themes. These themes were compared to those identified within the published literature. The results were intended to be indicative rather than comprehensive. They were used as a ‘sense’ or ‘temperature’ check of the findings from the published literature. Details of the studies’ characteristics are briefly outlined, but their synthesis is not reported in detail in the body of this thesis. However, a list of included studies, and a table of characteristics can be found in Appendices 9.7 and 9.8.

2.4 Details and characteristics of included published studies

2.4.1 Included published studies from searches in 2016

The following published studies from searches in 2016 were included in this review.

[1] Argyle, E. & Kelly, T. (2015) ‘Implementing person-centred dementia care: a musical intervention’, *Working with Older People: Community Care Policy & Practice*, 19(2), pp. 77–8.

[2] Bannan, N. and Montgomery-Smith, C. (2008) ‘“Singing for the Brain”: reflections on the human capacity for music arising from a pilot study of group singing with Alzheimer’s patients’, *The Journal of the Royal Society for the Promotion of Health*, 128(2), pp. 73–78.

[3] Billington, J. *et al.* (2013) ‘A literature-based intervention for older people living with dementia’, *Perspectives in Public Health*, 133(3), pp. 165–173.

[4] Burnside, L. D. *et al.* (2017) ‘here:now - conceptual model of the impact of an experiential arts program on persons with dementia and their care partners’, *Dementia*, 16(1), pp. 29–45.

[5] Byrne, L. and MacKinlay, E. (2012) ‘Seeking meaning: making art and the experience of spirituality in dementia care’, *Journal of Religion, Spirituality & Aging*, 24(1–2), pp. 105–119.

[6] Camic, P. M., Baker, E. L. and Tischler, V. (2016) ‘Theorizing how art gallery interventions impact people with dementia and their caregivers’, *The Gerontologist*, 56(6), pp. 1033–1041. doi: 10.1093/geront/gnv063.

[7] Camic, P. M., Tischler, V. and Pearman, C. H. (2014) ‘Viewing and making art together: a multi-session art-gallery-based intervention for people with dementia and their carers’, *Aging & Mental Health*, 18(2), pp. 161–168.

[8] Camic, P. M., Williams, C. M. and Meeten, F. (2011) ‘Does a “Singing Together Group” improve the quality of life of people with a dementia and their carers? A pilot evaluation study’, *Dementia*, 12(2), pp. 157–176.

[9] Chen, H. Y. *et al.* (2016) ‘Effects of TimeSlips on Cornell Scale for Depression in Dementia scores of senile dementia patients’, *International Journal of Nursing Sciences*, 3(1), pp. 35–38.

[10] Clement, S. *et al.* (2012) ‘Short and longer term effects of musical intervention in severe Alzheimer’s disease’, *Music Perception: An Interdisciplinary Journal*, 29(5), pp. 533–542.

[11] Cooke, M. *et al.* (2010a) ‘A randomized controlled trial exploring the effect of music on agitated behaviours and anxiety in older people with dementia’, *Aging & Mental Health*, 14(8), pp. 905–916.

[12] Cooke, M. *et al.* (2010b) ‘A randomized controlled trial exploring the effect of music on quality of life and depression in older people with dementia’, *Journal of Health Psychology*, 15(5), pp. 765–776.

[13] Davidson, J. and Almeida, R. (2014) ‘An exploratory study of the impact of group singing activities on lucidity, energy, focus, mood and relaxation for

persons with dementia and their caregivers', *Psychology of Well-Being*, 4(1), p. 24.

[14] Davidson, J. W. and Fedele, J. (2011) 'Investigating group singing activity with people with dementia and their caregivers: problems and positive prospects', *Musicae Scientiae*, 15(3), pp. 402–422.

[15] van Dijk, A. M., Van-Weert, J. C. M. and Dröes, R.-M. (2012) 'Does theatre improve the quality of life of people with dementia?', *International Psychogeriatrics*, 24(3), pp. 367–81.

[16] van Haeften-Van Dijk, A. M., Van Weert, J. C. M. and Droës, R. M. (2015) 'Implementing living room theatre activities for people with dementia on nursing home wards: a process evaluation study', *Aging & Mental Health*, 19(6), pp. 536–547.

[17] Dupuis, S. L. *et al.* (2016) 'Re-claiming citizenship through the arts', *Dementia*, 15(3), pp. 358–380.

[18] Eekelaar, C., Camic, P. M. and Springham, N. (2012) 'Art galleries, episodic memory and verbal fluency in dementia: an exploratory study', *Psychology of Aesthetics, Creativity, and the Arts*, 6(3), pp. 262–272.

[19] Flatt, J. D. *et al.* (2015) 'Subjective experiences of an art museum engagement activity for persons with early-stage Alzheimer's disease and their family caregivers', *American Journal of Alzheimer's Disease and Other Dementias*, 30(4), pp. 380–389.

[20] Fletcher, T. S. and Eckberg, J. D. (2014) 'The effects of creative reminiscing on individuals with dementia and their caregivers: a pilot study', *Physical & Occupational Therapy in Geriatrics*, 32(1), pp. 68–84.

[21] Fritsch, T. *et al.* (2009) 'Impact of TimeSlips, a creative expression intervention program, on nursing home residents with dementia and their caregivers', *The Gerontologist*, 49(1), pp. 117–127.

[22] George, D. R. and Houser, W. S. (2014) '"I'm a storyteller!": exploring the benefits of TimeSlips creative expression program at a nursing home', *American Journal of Alzheimer's Disease & Other Dementias*, 29(8), pp. 678–84.

[23] Gregory, H. (2011) 'Using poetry to improve the quality of life and care for people with dementia: a qualitative analysis of the Try to Remember programme', *Arts & Health*, 3(2), pp. 160–172.

[24] Gross McAdam, J. (2013) 'Conducting art-based research in dementia-specific healthcare in Australia', *UNESCO Observatory Journal in the Arts*, 3(3), pp. 1–20.

[25] Gross, S. M. *et al.* (2015) 'Creativity and dementia: does artistic activity affect well-being beyond the art class?', *Dementia*, 14(1), pp. 27–46.

[26] Guzmán-García, A., Mukaetova-Ladinska, E. and James, I. (2013) ‘Introducing a Latin ballroom dance class to people with dementia living in care homes, benefits and concerns: a pilot study’, *Dementia*, 12(5), pp. 523–535.

[27] Hara, M. (2011) ‘Music in dementia care: increased understanding through mixed research methods’, *Music and Arts in Action*, 3(2), pp. 34–58.

[28] Holm, A., Lepp, M. and Ringsberg, K. (2005) ‘Dementia: involving patients in storytelling - a caring intervention. A pilot study’, *Journal of Clinical Nursing*, 14(2), pp. 256–263.

[29] Holmes, C. *et al.* (2006) ‘Keep music live: music and the alleviation of apathy in dementia subjects’, *International Psychogeriatrics*, 18(4), pp. 623–630.

[30] Houser, W. S., George, D. R. and Chinchilli, V. M. (2014) ‘Impact of TimeSlips creative expression program on behavioral symptoms and psychotropic medication use in persons with dementia in long-term care: a cluster-randomized pilot study’, *American Journal of Geriatric Psychiatry*, 22(4), pp. 337–340.

[31] Kinney, J. M. and Rentz, C. A. (2005) ‘Observed well-being among individuals with dementia: Memories in the Making, an art program, versus other structured activity’, *American Journal of Alzheimer’s Disease & Other Dementias*, 20(4), pp. 220–227.

[32] Kontos, P. *et al.* (2016) ‘Elder-clowning in long-term dementia care: results of a pilot study’, *Journal of the American Geriatrics Society*, 64(2), pp. 347–353.

[33] Loizeau, A., Kündig, Y. and Oppikofer, S. (2015) ‘“Awakened Art Stories” - rediscovering pictures by persons living with dementia utilising TimeSlips: a pilot study’, *Geriatric Mental Health Care*, 3(2), pp. 13–20.

[34] MacPherson, S. *et al.* (2009) ‘An art gallery access programme for people with dementia: “you do it for the moment”’, *Aging & Mental Health*, 13(5), pp. 744-752 9p.

[35] Maguire, L. E. *et al.* (2015) ‘Participation in active singing leads to cognitive improvements in individuals with dementia’, *Journal of the American Geriatrics Society*, 63(4), pp. 815–816.

[36] McCabe, L., Greasley-Adams, C. and Goodson, K. (2015) ‘“What I want to do is get half a dozen of them and go and see Simon Cowell”: reflecting on participation and outcomes for people with dementia taking part in a creative musical project’, *Dementia*, 14(6), pp. 734–750.

[37] McDermott, O., Orrell, M. and Ridder, H. M. (2014) ‘The importance of music for people with dementia: the perspectives of people with dementia, family carers, staff and music therapists’, *Aging & Mental Health*, 18(6), pp. 706-716 11p.

[38] McIntyre, M. and Cole, A. (2008) ‘Love stories about caregiving and Alzheimer’s disease: A performative methodology’, *Journal of Health Psychology*, 13(2), pp. 213–25.

[39] Musella, O. *et al.* (2009) ‘Visual art improves communication and behaviour of AD patients’, in *New Trends in Alzheimers and Parkinson Related Disorders: ADPD 2009*, pp. 15–20.

[40] Narme, P. *et al.* (2014) ‘Efficacy of musical interventions in dementia: evidence from a randomized controlled trial’, *Journal of Alzheimer’s Disease*, 38(2), pp. 359–369.

[41] Osman, S. E., Tischler, V. and Schneider, J. (2016) ‘“Singing for the Brain”: a qualitative study exploring the health and well-being benefits of singing for people with dementia and their carers’, *Dementia*, 15(6), pp. 1326–1339.

[42] Palo-Bengtsson, L., Winblad, B. and Ekman, S. L. (1998) ‘Social dancing: a way to support intellectual, emotional and motor functions in persons with dementia’, *Journal of Psychiatric and Mental Health Nursing*, 5, pp. 545–554.

[43] Palo-Bengtsson, L. and Ekman, S.-L. (2000) ‘Dance events as a caregiver intervention for persons with dementia.’, *Nursing Inquiry*, 7(3), pp. 156–165.

[44] Palo-Bengtsson, L. and Ekman, S. (2002) ‘Emotional response to social dancing and walks in persons with dementia’, *American Journal of Alzheimer’s Disease & Other Dementias*, 17(3), pp. 149–153.

[45] Petrescu, I., MacFarlane, K. and Ranzijn, R. (2014) ‘Psychological effects of poetry workshops with people with early stage dementia: an exploratory study’, *Dementia*, 13(2), pp. 207–215.

[46] Phillips, L. J., Reid-Arndt, S. A. and Pak, Y. (2010) ‘Effects of a creative expression intervention on emotions, communication, and quality of life in persons with dementia’, *Nursing Research*, 59(6), pp. 417–425.

[47] Rentz, C. A. (2002) ‘Memories in the Making: outcome-based evaluation of an art program for individuals with dementing illnesses’, *American Journal of Alzheimer’s Disease & Other Dementias*, 17(3), pp. 175–181.

[48] Roe, B. *et al.* (2016) ‘Coffee, cake & culture: evaluation of an art for health programme for older people in the community’, *Dementia*, 15(4), pp. 539–559.

[49] Sakamoto, M., Ando, H. and Tsutou, A. (2013) ‘Comparing the effects of different individualized music interventions for elderly individuals with severe dementia’, *International Psychogeriatrics*, 25(5), pp. 775–784.

[50] Särkämö, T. *et al.* (2014) ‘Cognitive, emotional, and social benefits of regular musical activities in early dementia: randomized controlled study’, *The Gerontologist*, 54(4), pp. 634–650.

[51] Sauer, P. E. *et al.* (2016) ‘“It makes me feel like myself”: person-centered versus traditional visual arts activities for people with dementia’, *Dementia*, 15(5), pp. 895–912.

[52] Sixsmith, A. and Gibson, G. (2007) ‘Music and the wellbeing of people with dementia’, *Ageing and Society*, 27(01), pp. 127–145.

[53] Stevens, J. (2012) ‘Stand up for dementia: performance, improvisation and stand up comedy as therapy for people with dementia; a qualitative study’, *Dementia*, 11(1), pp. 61–73.

[54] Sung, H. *et al.* (2006) ‘The effects of group music with movement intervention on agitated behaviours of institutionalized elders with dementia in Taiwan’, *Complementary Therapies in Medicine*. 14(2), pp. 113–119.

[55] Sung, H. C. *et al.* (2012) ‘A group music intervention using percussion instruments with familiar music to reduce anxiety and agitation of institutionalized older adults with dementia’, *International Journal of Geriatric Psychiatry*, 27(6), pp. 621–627.

[56] Swinnen, A. M. C. (2016) ‘Healing words: a study of poetry interventions in dementia care’, *Dementia*, 15(6), pp. 1377–1404.

[57] Ullán, A. M. *et al.* (2013) ‘Contributions of an artistic educational program for older people with early dementia: an exploratory qualitative study’, *Dementia*, 12(4), pp. 425–446.

[58] Unadkat, S., Camic, P. M. and Vella-Burrows, T. (2017) ‘Understanding the experience of group singing for couples where one partner has a diagnosis of dementia’, *Gerontologist*, 57(3) (Online first 2016), pp. 469–478.

[59] Ward, A. R. and Parkes, J. (2017) ‘An evaluation of a Singing for the Brain pilot with people with a learning disability and memory problems or a dementia.’, *Dementia.* 16(3) (Online first 2015), pp. 360–374.

[60] Young, R. *et al.* (2015) ‘The impact of viewing and making art on verbal fluency and memory in people with dementia in an art gallery setting’, *Psychology of Aesthetics, Creativity, and the Arts*, 9(4), pp. 368–375.

2.4.2 Included published studies from update in 2018

The following published studies from the update search in 2018 were included in this review.

[61] Basting, A. (2018) ‘Building creative communities of care: arts, dementia, and hope in the United States.’, *Dementia*, 17(6), pp. 744–754.

[62] Belver, M. et al. (2018) ‘Art museums as a source of well-being for people with dementia: an experience in the Prado Museum’, *Arts and Health*, 10(3), pp.213-226.

[63] Boersma, P. et al. (2017) ‘Implementation of the Veder contact method in daily nursing home care for people with dementia: a process analysis according to the RE-AIM framework’, *Journal of Clinical Nursing*, 26(3–4), pp. 436–455.

[64] Broome, E., Dening, T. & Hill, O. (2018) 'Participatory arts in care settings: a multiple case study: innovative practice', *Dementia*, (online first 24 October 2018).

[65] Broome, E., Dening, T. and Schneider, J. (2017) 'Facilitating Imagine Arts in residential care homes: the artists' perspectives', *Arts & Health*, (online first December 2017).

[66] Campbell, S. et al. (2017) '"Music in Mind" and Manchester Camerata: an exploratory qualitative evaluation of engagement in one care home in northwest England', *Quality in Ageing & Older Adults*, 18(1), pp. 69–80.

[67] Daykin, N. et al. (2018) 'The role of participatory music making in supporting people with dementia in hospital environments', *Dementia*, 17(6), pp. 686–701.

[68] Eades, M., Lord, K. and Cooper, C. (2016) '"Festival in a Box": Development and qualitative evaluation of an outreach programme to engage socially isolated people with dementia', *Dementia*, (online first August 2016).

[69] Evans, S. C., Garabedian, C. and Bray, J. (2019) '"Now he sings". The My Musical Memories reminiscence programme: personalised interactive reminiscence sessions for people living with dementia', Dementia, 18(3), pp. 1181–1198.

[70] Hendriks, I. et al., (2018) 'Implementation and impact of unforgettable: an interactive art program for people with dementia and their caregivers', *International Psychogeriatrics* (online first 18 July 2018).

[71] Johnson, J. et al. (2017) 'Museum activities in dementia care: using visual analog scales to measure subjective wellbeing, *Dementia*, 16(5), pp 591-610.

[72] Keating, F., Cole, L. and Grant, R. (2018) 'An evaluation of group reminiscence arts sessions for people with dementia living in care homes', *Dementia*, (online first 16 July 2018).

[73] Kontos, P. et al. (2017) 'Presence redefined: the reciprocal nature of engagement between elder-clowns and persons with dementia', *Dementia*, 16(1), pp. 46–66.

[74] Longden, E. et al. (2016) 'An evaluation of shared reading groups for adults living with dementia: preliminary findings', *Journal of Public Mental Health*, 15(2), pp. 75–82.

[75] Low, L. F. et al. (2016) 'We think you can dance! a pilot randomised controlled trial of dance for nursing home residents with moderate to severe dementia', *Complementary Therapies in Medicine*, 29, pp. 42–44.

[76] Mc Parland, P., Cutler, C. and Innes, A. (2017) 'The challenges of implementing and evaluating a pilot music and movement intervention for people with dementia: Innovative practice', *Dementia*, 16(8), pp. 1064–1068.

[77] Mittelman, M. S. and Papayannopoulou, P. M. (2018) 'The Unforgettables: a chorus for people with dementia with their family members and friends', *International Psychogeriatrics*, 30(6), pp. 779–789.

[78] Mondro, A. et al. (2018) 'Retaining identity: creativity and caregiving', *Dementia*, (online first 4 October 2018).

[79] Newman, A. et al. (2019) 'The role of the visual arts in the resilience of people living with dementia in care homes', *Ageing and Society*, 39(11), pp. 2465–2482.

[80] Schall, A. et al. (2018) 'Art museum-based intervention to promote emotional well-being and improve quality of life in people with dementia: the ARTEMIS project', *Dementia*, 17(6), pp. 728–743.

[81] Swinnen, A. and de Medeiros, K. (2018) '"Play" and people living with dementia: a humanities-based inquiry of TimeSlips and the Alzheimer's Poetry Project', *Gerontologist*, 58(2), pp. 261–269.

[82] Tan, M. (2018) 'Connecting reminiscence, art making and cultural heritage: a pilot art-for-dementia care programme', *Journal of Applied Arts & Health*, 9(1), pp. 25-36.

[83] Tyack, C. et al. (2017) 'Viewing art on a tablet computer: a well-being intervention for people with dementia and their caregivers', *Journal of Applied Gerontology*, 36(7), pp. 864–894.

[84] Windle, G. et al. (2018) 'The impact of a visual arts program on quality of life, communication, and well-being of people living with dementia: a mixed-methods longitudinal investigation', *International Psychogeriatrics*, 30(3), pp. 409–423. 49

2.4.3 Characteristics of included published studies

The studies explored activities including participatory music-making, group singing, shared reading, art-viewing/discussion/education and art-making in gallery and museum settings, visual arts, creative expression (art-based storytelling), living room theatre, creative reminiscence, poetry, dance, elder clowning, opera, intergenerational arts activity, stand-up comedy, group music and movement, and art-based activity delivered using a touchscreen device. Most activities took place in residential and nursing settings, with others delivered in day centres, community settings or cultural or heritage sites such as galleries and museums. A much smaller number took place in hospitals and participants' own homes. All levels of dementia were represented, and studies included activities undertaken by people with dementia on their own and with professional or family caregivers. The majority of studies featured activities taking place in the UK.

Other studies were conducted in Australia, Canada, China, Finland, France, Germany, Republic of Ireland, Italy, Japan, Netherlands, Singapore, Spain, Sweden, Switzerland, Taiwan, and USA.

Studies addressed a wide range of aims. For most, these comprised measurement or observation of the effects of activities on the common behavioural and psychological symptoms of dementia, cognitive function, subjective and observed wellbeing, quality of life, communication and engagement, activities of daily living, and carer ‘burden’. Studies also explored the qualitative experience of those taking part in activities; these addressed the role of the arts in participants’ lives, their responses to activities in terms of their suitability or acceptability, and in relation to interaction and engagement with others (including family and professional caregivers). A smaller number addressed questions directly relevant to implementation and facilitation of arts activities.

A table of characteristics can be found in Appendix 9.3.

2.4.3.1 A typology of included studies

Because my interest in this review was in understanding the limitations of studies and the challenges faced by authors, it is first useful to describe top-level characteristics of the included studies in relation to their design. This is because limitations of particular study designs and methods are well-understood, at least in the academic context.

To understand the nature of the quantitative and mixed methods studies included, I have referred to the definitions and descriptions given in the Cochrane Handbook for Systematic Reviews of Interventions, and in the Cochrane Community Glossary online. However, many study authors do not identify their study design and the majority of studies do not easily fit definitions of what constitutes rigorous experimental or quasi-experimental study from either Cochrane or from social research disciplines (e.g. Alasuutari, Bickman and Brannen, 2008). The majority of quantitative and mixed methods studies lack randomisation or equivalent matching of participants or groups, and most also lack suitable comparison or control. Very few authors attempted concealed allocation or the blinding of assessors or participants. Many studies in this field, if quality assessment were to be applied, might be considered ‘weak’ or ‘moderate’.

However, as noted earlier, I have not applied quality assessment standards to the literature in this review. In understanding the nature of the literature included, I found it useful to turn away from the idea of a hierarchy of evidence in describing its characteristics, and towards that of a *typology* in which evidence is categorised according to both its design and the nature of the research question driving it (see Figure 1 below, an example borrowed from Muir Gray and cited in Petticrew and Roberts, 2003, p. 528).

Figure 1: Example typology of evidence

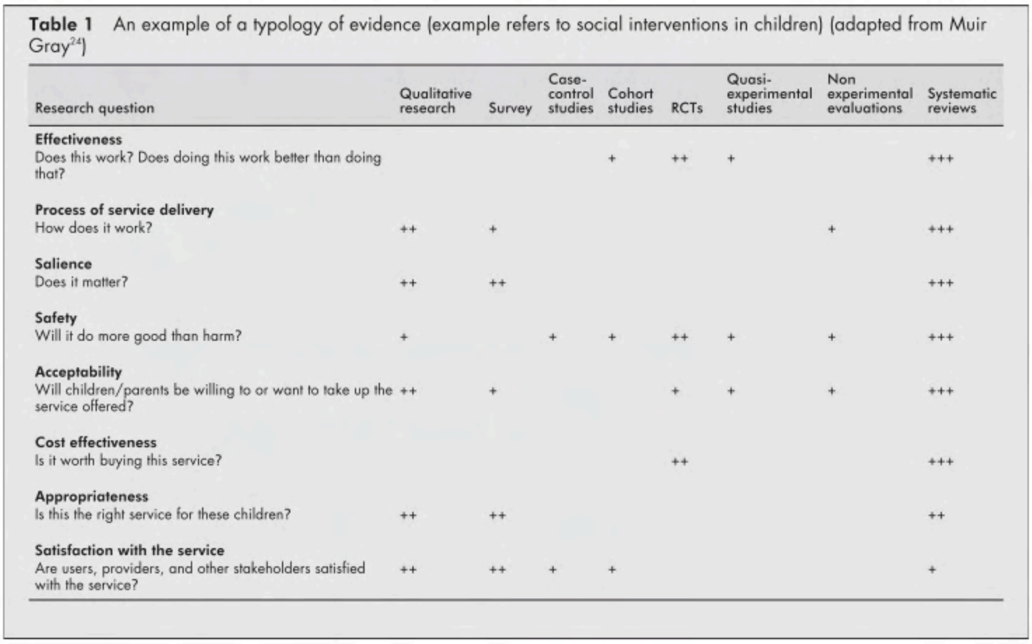


Table 1 An example of a typology of evidence (example refers to social interventions in children) (adapted from Muir Gray[24])

Research question	Qualitative research	Survey	Case-control studies	Cohort studies	RCTs	Quasi-experimental studies	Non experimental evaluations	Systematic reviews
Effectiveness Does this work? Does doing this work better than doing that?				+	++	+		+++
Process of service delivery How does it work?	++	+					+	+++
Salience Does it matter?	++	++						+++
Safety Will it do more good than harm?	+		+	+	++	+	+	+++
Acceptability Will children/parents be willing to or want to take up the service offered?	++	+			+	+	+	+++
Cost effectiveness Is it worth buying this service?					++			+++
Appropriateness Is this the right service for these children?	++	++						++
Satisfaction with the service Are users, providers, and other stakeholders satisfied with the service?	++	++	+	+				+

Typologies of this kind provide a useful means through which to understand and conceptualise the strengths and weaknesses of different methodological approaches.

The following table outlines the characteristics of the published studies included in this review, employing categories adapted from those in the above example.

Table 1: Included studies characterised by study type

Study category	**Included studies, identified according to numbered list above**
Qualitative only	1, 4, 5, 8, 16, 17, 22, 23, 26, 27, 28, 36, 37, 38, 41, 43, 44, 45, 48, 52, 53, 56, 57, 58, 59, 62, 63, 64, 65, 66, 73, 76, 76, 78, 81, 82
Quantitative only	9, 10, 11, 12, 15, 21, 29, 30, 31, 32, 35, 40, 46, 47, 49, 51, 54, 55, 60, 71, 72, 74, 75, 80
Mixed methods	2, 3, 6, 7, 13, 14 18, 19, 20, 24, 25, 33, 34, 39, 42, 50, 61, 67, 69, 70, 77, 79, 83, 84
Randomised controlled study	10, 11, 12, 21, 29, 30, 40, 49, 50, 54, 55, 75, 80
Quasi-experimental (comparative/controlled study	15, 20, 46, 71
Cohort study	84
Randomised, no control	21
Case study	63, 65, 70
Process evaluation or implementation study	1, 16, 63, 64, 70, 78
Non-experimental evaluation research study	2, 3, 6, 7, 9, 13, 14, 18, 19, 25, 26, 31, 32, 33, 34, 35, 39, 42, 47, 51, 60, 61, 67, 69, 72, 74, 77, 83

As Table 1 shows, of the studies included, 36 were qualitative, 24 used mixed methods and 24 were solely quantitative in method. There were no studies that might truly be classified as falling into the categories of case control or cross-sectional survey. One can be categorised as a prospective cohort study. There were no economic evaluations.

Quasi-experimental studies are those that, through the way in which subjects are selected, and their purpose, are intended to mimic randomised experiments (Cook and Wong, 2008). Although some of the studies categorised as ‘non-experimental evaluations’ may have included comparison or control groups, it was not clear from reporting that they had otherwise taken steps, either through design or analysis, to mitigate for the lack of randomisation in their comparison.

Because all studies relied on the introduction of an ‘intervention’, rather than observation of a cohort of individuals over a period of time, I have not used the category of ‘time series’, although it is noted below that a number of studies did use repeated measures over the course of a programme. Only those studies termed by their authors as ‘case studies’ have been listed as such, although it could be argued that a good number of the other included studies could perhaps fit this categorisation. Those studies noted as process evaluations or implementation studies are of varying type, with only one using a defined process evaluation framework.[16] Others focused, for example, on identifying barriers and facilitators to arts activities in the settings in which they were delivered.

The category of non-experimental evaluation research study (n=28) is deliberately broad. This reflects the inclusion of those quantitative or mixed methods studies that attempt, for example, through observation or repeated measures during an activity, or pre and post measurement, to assess the effects of that activity on its subjects. In these studies, the 'intervention' is not under the control of the evaluator, the authors do not use random assignment, and while some do include a comparative element, most do not attempt any rigorous comparative analysis. While some purely qualitative studies self-identify as 'evaluations' or describe 'evaluation' as a key aim, they are not included in this category. [1, 23, 36, 48, 53, 59, 62, 66, 68, 76]

A characterisation of studies using a typology of this kind, rather than through an assessment of quality according to a given hierarchy, opens up a space for commentary upon the relationship between the design and methods of studies and their aims or research questions. It suggests that non-experimental evaluations of the kind described above, as well as quasi-experimental evaluations, may be poorly suited to producing 'strong' evidence in response to many kinds of question. In contrast, qualitative and other kinds of studies have value for providing answers of widely applicable relevance to questions around process (how does it work?), salience (does it matter?), acceptability (will the subjects take it up?), appropriateness (is this the right kind of thing to be offering?), and satisfaction (are stakeholders satisfied with the service?) – even if they are not appropriate for demonstrating that a service has been effective.

It is noteworthy, however, that the majority of studies in this review, including some of those using qualitative methods, were attempting to answer questions relating to the effectiveness of the activities involved.

2.4.3.2 Other characteristics – methods and approaches

The landscape of published research in this field includes further interesting features. Seventeen included studies defined themselves as ‘pilots’ or ‘exploratory’, and 18 self-identified as ‘evaluation’. In only 16 studies had follow-up research (classified here as four weeks or more post intervention) been attempted. Repeated measures throughout a programme of activity were used in 20 studies; pre and post programme measurement was used in 17. In 39 studies, authors had applied measures or used observational methods for collection of data during activity sessions themselves. When using observational methods, seven studies involved participant observation, and in 25 non-participant observational tools and measures had been applied. In 36 studies, authors noted the use of either self-report survey measures with people living with dementia, or interviews intended to elicit their perspectives and opinions first-hand.

For those authors using qualitative methods, interview or focus groups were the most popular method used to collect data. There were five ethnographic studies, five that used an arts-based or creative research method, two that applied a phenomenological approach, and two that used participatory methods. Either video or audio recording (not including audio-recording of interviews) was used in 24 studies to collect data. Artist reflective journals or other project documentation were used as data in 10 studies.

2.4.4 Characteristics of grey literature studies

All grey literature studies or reports consulted were from the UK. Eight were doctoral level dissertations and the remainder identified themselves as evaluations. Seven were qualitative, 19 used mixed methods, and three were solely quantitative in approach.

Repeated measures over different sessions had been applied in 11 studies; observations or measures during sessions in 15, and; 17 used interview or focus groups. Some form of first-hand accounts from people living with dementia were included in 14 of the included items. Artist reflective journals were used as data in eight, and nine used video or audio recordings.

The stated aim of most studies or reports was to evaluate the impacts or effects of particular programmes of activity on the wellbeing or quality of life of individuals with dementia. Additionally, there was a greater focus than in published studies on the development and facilitation of the activities and on organisational impacts and outcomes. All the evaluation reports had been commissioned or were conducted by the organisations delivering the evaluated programme. The dissertations included discussion of limitations and reflexivity, but many of the evaluation reports gave little detail on these, with some including only very brief accounts of how data were collected or analysed. However, a number of the studies included content that, in a different and peer-reviewed form, also appeared within the published studies. (See Appendices 9.5 and 9.6 for a full list of included studies and a table of characteristics.)

2.5 Synthesis of included published studies

Three themes and their sub-themes are reported in this section. These are:

1. Designing studies
2. Measuring change
3. People and settings

2.5.1 Theme 1: Designing studies

Evaluators and researchers attempting experimental designs reported difficulties with all aspects of the process, including randomisation, blinding, and the use of comparator or control groups in complex natural settings. Authors acknowledged the variables or potential variables involved in studies, noting how they added complication to analysis or presented barriers to confident theorisation. Issues related to the delivery of activities, and the inability of evaluators to control this was described (e.g. how much, where, how often, and the barriers and facilitators or mediators for effects). While qualitative research studies were common, in mixed methods approaches, qualitative methods were often seen as augmenting quantitative evaluation findings about effectiveness, rather than contributing understanding in response to alternative kinds of evaluation question.

Many study authors commented on some very commonly understood limitations resulting from the way in which studies were designed or conducted. For example: difficulties in generalising from study findings, or the risk that, through systematic error or bias, studies may have over-estimated or under-estimated effects of the activities or programmes.

To provide a point of reference for this, in contrast to bias, *external validity* may be defined as the extent to which the results of a study might be applicable to other populations or contexts. *Internal validity* refers to the extent to which results can be considered applicable or true for the particular population or setting involved. As an example of how risk of bias might be assessed, for randomised trials Cochrane recognises risks that can result from four main domains of bias if they are not mitigated through design or statistical methods (Higgins *et al.*, 2019, section 8.2.2). These are bias: (1) arising from the randomisation process; (2) due to deviations in the intended interventions (including those that result from failure to blind); (3) due to missing data; (4) in measurement of the outcome (because of use of inappropriate measures, for example, or as a result again of lack of blinding of assessors); (4) in selection of the reported result. Included studies reported limitations resulting from all four kinds of bias.

The most salient methodological features relevant to reported limitations and challenges in the studies included in this review are listed in Table 2. They will be further described and discussed below.

Table 2: Theme 1: Designing studies

Theme	**Sub-themes**	**Description**
Designing studies	Randomisation Blinding Comparator or control groups Variables Sample size Longitudinal and follow up Attendance and attrition (missing data) Subjectivity Use of qualitative methods	Authors reported difficulties with many of the key features of experimental and quasi-experimental designs in evaluating and researching arts and dementia. These included randomisation, blinding, and comparative analysis. Sample size was frequently small, and attendance and attrition caused problems. There were few longitudinal studies and no included economic evaluations. Qualitative approaches were common and were sometimes valued as augmenting or interpreting quantitative findings.

2.5.1.1 Randomisation

Convenience or self-selected samples were a common feature among all studies, and in the included non-experimental evaluation research studies in particular. Authors not using randomised samples commonly and correctly report that observed differences in outcomes cannot be confidently attributed to an activity because of potential biases in the selection and assignment of participants to an intervention group.

Selection biases affecting samples mean that study results may not be representative of the dementia population as a whole. Examples referenced by authors include the selection of participants by staff in a care setting based on a sense that certain of them will benefit or enjoy an activity, [3, 26, 42, 72] or the inclusion in a study of only those with capacity to consent. [66] Many evaluated projects were delivered in residential care settings or nursing homes, with the result that age ranges, the prevalence of particular dementia stages, and gender balance may have been skewed.

Few authors discuss their reasons for not randomising, or difficulties with it, although when they do these can be revealing. In one study, authors cited ethical sensitivity to the cultural environment of a long-term care facility whose management was unwilling to allow some residents to receive an intergenerational arts activity programme while other residents did not. [51] Authors of another study comparing results for residents who were receiving a creative expression activity to those in the recruited facilities not receiving it, noted that random assignment would have led to wide disparities in group sizes. [46]

2.5.1.2 Blinding

Blinding of participants is understood to be very difficult, if not impossible, in social research studies taking place in natural settings. As noted above, a failure to blind those involved in receiving or conducting a study may have contributed to biases in the way in which arts activities or programmes were implemented and their effects evaluated or reported. A small number of included studies attempted blinding of assessors. [11, 12, 15, 29, 40, 49, 50] Of these, one study involving a comparison between a music and a cooking activity [10] repeated the intervention with blinded assessors evaluating participants using videoed interviews and reported significantly different (less positive) results in the second iteration. [40]

2.5.1.3 Comparison or control groups

A lack of comparison or control groups is recognised as hampering the confident attribution of causality to an intervention, as well as the uncovering of its mechanisms of impact. Identification and implementation of a comparison activity were reported to be challenging in a number of studies. [11, 12, 13, 40, 55, 51, 75, 83] This included the comparison of non-standard activities to a 'usual care' control group. In particular, it was noted that 'usual care' might be facilitated differently across multiple care settings in ways that the evaluator could not predict, including the fact that a range of alternative meaningful activities were available to residents, [72] or that the introduction of an intervention within a single setting might itself have ripple effects for those individuals not receiving it. [55] The difficulties of matching comparator groups within different care settings because of differences resulting from variables such as setting size, care philosophy, and location were discussed in some studies attempting this. [15, 21] Outside the closed environment of a clinical or care setting, establishing control conditions or 'usual care' may be even more problematic.

The mechanisms effecting change for a given activity or programme were not always well defined at the start of a study, and this sometimes made developing or identifying a suitable comparison activity hard. In two studies reporting on a cross-over within subjects trial assessing the effects of a music intervention and a reading control group on depression and quality of life in groups of care home residents with dementia, the two activities were very different in structure and delivery. [11, 12] These differences were said to have contributed to problems in interpreting data and there was little indication that significant improvements for both groups were the result of anything other than engagement in arts activity as opposed to routine care. Another study looked at the effects of an intervention involving a person-centred intergenerational visual arts program in which a volunteer interacted with people with dementia on an individual basis. [51] A sub-sample of study participants was also observed engaging in traditional visual arts activities. However, this comparison activity did not include the 1:1 component of the intergenerational intervention, and it was therefore deemed difficult to separate out the effects of increased personal interaction from those related to the arts.

2.5.1.4 Variables

Authors acknowledged the many variables or potential variables involved in studies, noting how an inability to account for them added complication to analysis of data, affected claims to generalisability, or presented barriers to confident theorisation around causality. Authors sometimes simply used a catch-all statement that confounding variables may have affected findings.

Where samples were not randomised, a reported observed or potential imbalance in socio-demographic characteristics, including prior interest or experience of the arts, [4, 8, 60, 71, 80] ethnicity, [46, 22, 41, 84] and gender [30, 71, 22] were sometimes cited as factors limiting the confidence authors had in the generalisability of a study's results. The effect of sexual orientation was not explored, although the authors of one study noted that it may play a role where an activity's mechanisms involve a couple relationship component. [58] Included studies often did not report or record data of these kinds, sometimes stating as much. [22, 40, 47, 51, 74] Such decisions will have limited the data analysis that was possible, as the authors of a study exploring the effects of participatory music sessions in hospitals noted, [67] or necessitated robust analysis to mitigate for variables, which was not always conducted.

Functional or clinical impairments, co-morbidities, and medication use were also often not accounted for, or sometimes reported as out of scope. [32, 51, 67] Many studies did not differentiate, or differentiated only very broadly between dementia types or level of severity of the condition among their participants, with some noting the absence of a clinical diagnosis of dementia. This means that there is often a lack of clarity as regards how applicable results can be to individuals from different groups.

Authors described difficulties in designing programmes of arts activities that offered an equivalent experience at each iteration. While some authors developed a procedures manual, [11, 12] or produced standardised guidelines for intervention delivery, [84] some studies relied upon use of proprietary or franchisable programmes such as the trademarked TimeSlips, [3, 21, 22, 30, 35, 46, 81] Singing for the Brain, [3, 27, 41, 59] or shared reading sessions delivered by the Reader Organisation. [3, 74]

Some researchers who were also artist practitioners explored their practice through their research, developing protocols for delivering sessions based upon this experience. [24, 26] However, the inherent flexibility and unpredictability that forms a part of creative facilitation and people's responses to it, remains one of the key variables for which evaluators found it difficult to account. A requirement to force an intervention to 'fit', regardless of organisational routines, facilitator styles and individual participant needs, was recognised by some authors as being a potential limiter of its effects. [12, 49]

2.5.1.5 Sample size

Sample size will have repercussions for calculation of power and effect size, statistical significance, and the subsequent validity of claims a study author can make, including for the generalisability of results and attribution of causality (Hancock, 2004).

In the studies reviewed, arts-based activities were most often experienced in small groups or by individuals. As a result, sample sizes were usually small. Small sample sizes were very commonly reported as contributing to authors' inability to generalise from the findings, [7, 10, 13, 19, 20, 30, 31, 32, 33, 40, 49, 54, 74, 77, 82] including in qualitative studies. [5, 26, 45, 48, 53, 57] For authors wishing to establish the effectiveness of arts activities, small participant numbers caused particular difficulties, including limiting choices of design and the kinds of measures and tools that could be used, as well as their accuracy. However, one author made the practical point that, in a mixed methods study, smaller numbers had made the collection and analysis of qualitative data manageable. [83] In another study, single case methodology was recommended for its capacity to enable individualised analyses of efficacy. [26]

2.5.1.6 Longitudinal design and follow-up

Longitudinal designs were uncommon. Only fifteen studies attempted to follow up with participants four weeks or longer after an intervention. Most of these follow-ups were inconclusive or showed no evidence of long-term effects. Only two studies suggested the potential for significant and sustained benefits for participants. [50, 84] A lack of information about long-term effects of arts activities was stated as a limitation in relation to the findings of multiple studies. [3, 15, 20, 21, 22, 29, 31, 49, 54, 57, 58, 62, 72, 66] Because many studies were of short term programmes of activity, some authors suggested that the length of study may have been insufficient for any positive changes for individuals or groups to be detectable, [6, 14, 18, 21] as well as for indirect change to have been effected within complex care environments. [21, 22] Time, funding, and resource restrictions may have affected the design and conduct of some studies. One author, for example, stated clearly the potentially negative effects of having to work with a short intervention because of funding and scheduling issues relating to its implementation. [50]

2.5.1.7 Attrition and attendance (missing data)

Problems with recruitment and attrition were multiple and widespread and this review of the literature showed how challenging it can be to recruit sufficient, stable or appropriate participant groups in this field. Reported reasons for attrition or non-attendance at sessions included: participant choice; frailty, illness, infirmities and co-morbidities relating to aging; symptoms related to the natural progression of the condition, and; mortality. Evaluators therefore reported gaps in data, decreased numbers in participant groups, and other knock-on effects such as unplanned variations in group sizes across sites. One study of the effect of guided creative reminiscence activities on the quality of life of clients with memory loss, for example, reported a fall from 36 participants interested in taking part, to just 12 completing the four weekly hour-long sessions assessed. [20] Following up participants in the community could also be problematic. [6] In contrast, one large scale study reported few problems, but the authors did detail a robust recruitment strategy that may have mitigated for the difficulties. [84]

Missing data introduce the risk of bias in experimental studies, if not accounted for in analysis. The simplification, aggregation, or collapsing of complex observational data into more general categories was a potentially problematic mitigation approach taken in some studies. [31, 34, 51]

2.5.1.8 Subjectivity

Different approaches, depending upon the particular methodology adopted, were used to increase the objectivity of studies and the validity of their results.

Few authors referenced measures and checks of inter-rater reliability and the training and procedures required to maintain it. The problems that might result from this were illustrated in a study of a creative expression programme where raters had to be retrained partway through because disparities had become obvious. [46] In another it was suggested that novice observers might have become habituated to participants' behaviours over the course of a study, unconsciously amending their rating as a result. [25]

Qualitative researchers sometimes approached the issue of their own subjectivity through reflexive discussion, [e.g. 27, 43, 56] bracketing, [44, 83] through checking of analysis and findings with colleagues, or through some form of validation with participants (although this validation was seldom carried out with people with dementia).

Authors sometimes expressed concerns about research participation effects and response bias, including the possibility that subjects of a study might have altered their behaviour in response to awareness of being observed or questioned. In one study, care staff members delivering an activity reported feeling nervous as a result of being observed and because of the presence of cameras. [15] In another, the authors praised the unique perspective and opportunities offered by use of video, but discussed the possibility that participants' conduct might have been modified by its presence. [73] There were also concerns that care staff might have unconsciously or consciously changed their interpretations of an activity in order to please the researcher, [22] that artists' responses might have been shaped by the requirements of the arts organisation commissioning them, [64] or that a participant might have responded to a perceived demand of a study measuring quality of life to appear ‘normal’. [14]

2.5.1.9 Use of qualitative methods

In mixed methods studies, positive findings sometimes emerged from the qualitative component of the study when the quantitative data were inconclusive. [5, 7, 14, 83] They were otherwise reported as being useful in order to augment these data, to reveal nuance, or to provide a point of access to subjective individual experience or the voices and perspectives of participants. [4, 13, 84] The value of qualitative methods was signposted through descriptors of the derived data and analyses such as ‘richness’ and ‘depth’. The word ‘triangulation’ is occasionally used by authors to describe the use of a variety of methods or measures that enable understanding of a phenomenon from different perspectives. For example, authors might suggest that observational methods be used alongside interviews in future research [e.g. 36, 63] or that self-report measures would have complemented observation. [72] Some implied or stated that the use of mixed methods had enhanced the validity of a study, combatting potential problems caused by use of only one method or tool. This was also sometimes framed as a recommendation for future study.

Several authors suggested that an emphasis on technically correct application of designs and methods to measure the effects of the arts (its outcomes) came at the expense of understanding its importance or meaning for those engaging in activities, [31, 66, 81] or of the social and cultural contexts determining this involvement. [27, 69, 80]

Although authors of qualitative or mixed methods evaluation research studies did not always report limitations in the same way as their quantitative counterparts, many did. They referred, for example, to the way in which results could not be considered generalisable because of the absence of control conditions or the size of a sample. This perhaps raises questions about the criteria authors in this field are applying in their assessment of the validity of study findings and their choices regarding appropriateness of study design (Hammersley, 2008). And, while the literature suggested a widespread acceptance of the utility of mixed methods approaches, there was little discussion of potential issues resulting from the mixing of methods whose origins lie in paradigms with differing epistemological assumptions.

2.5.2 Theme 2: Measuring change

In section 2.5.1, I discussed limitations and challenges reported in the included studies that relate to study design. Many of the issues described would be familiar to those conducting any evaluative study of a social intervention taking place in a naturalistic setting. Although the challenges may be prevalent in the arts and dementia field, they are hardly unique to it. In this second thematic area, I present those issues that related to authors' attempts to understand the effects of arts activities upon participants, and to measure or 'value' any changes that might result. These themes are described in Table 3 and reported below.

Table 3: Theme 2: Measuring change

Theme	**Sub-themes**	**Description**
Measuring change	Use and adaptation of standardised measures and tools Baselines and ceiling or floor effects Measurement in the context of dementia Measurement in the context of arts activity	Evaluators found the use of standardised measures and tools problematic. For some this had led to adaptation or creation of new tools. Identification of change was problematic because of difficulties around baseline measurement and recruitment issues. There were disagreements about the measures and methods used to assess some of the key constructs. The effects of participation in arts activity sometimes appeared to resist measurement and interpretation.

2.5.2.1 Use and adaptation of standardised measures and tools

Validated standardised tools were used successfully in included studies, but there were also numerous reported issues with them. Most standardised self-report measures currently available for those with dementia are, of necessity, brief. They therefore may only provide limited information about the perspectives of people with dementia on subjective constructs such as quality of life or wellbeing. It has been described as critical to employ qualitative tools alongside quantitative measures in order to access these perspectives (Stewart-Archer *et al.*, 2016, p. 2). Such a view is supported by results from studies of arts activities in which standardised measures and qualitative tools were combined, or those in which a stand-alone standardised measure was described as insufficient to provide conclusively useful results. [e.g. 6, 7, 14, 83]

Some authors suggested that available tools lacked the specificity required to measure constructs of significance in the context of arts and dementia activity. [14, 19, 20, 79, 80] In a pilot evaluation of a singing group, inconclusive results from existing quality of life measures were linked to ongoing difficulties in adequately measuring the construct in those with a deteriorating condition, and their unreliability in the context of moderate or severe cognitive impairment in dementia. [6] While appropriate for capturing micro-changes in the day-to-day context of an individual's experience of person-centred care, DCM was found unsuitable for assessing an activity's long-term effects in one study. [72] The authors of another study raised the question of whether standardised measures might create standardised responses, and what that might mean for measurement of change resulting from an intervention whose facilitation invites non-standardised responses. [37] Authors of a further study reported participants questioning the validity of the measures being used, suggesting that for people with dementia in particular, wellbeing and quality of life are dynamic phenomena, varying from hour to hour for reasons often not related to the evaluated activity. [20] The selection and application of measures with the ubiquitous small sample sizes and short time frames of studies in the field was also described as problematic, contributing perhaps to problems in reliably detecting significant change. [7, 18]

Dissatisfaction with existing measures sometimes led authors to adapt or develop new ones, but whilst this may have led to greater specificity, authors also frequently acknowledged that it introduced limitations, particularly in relation to the generalisability of results, confidence in the validity of the construct being measured, and issues when several different variables were under consideration. For example, in a mixed methods study of a creative expression activity in an art museum, researchers used the two relevant subscales of the Caregiver Burden Inventory and adapted the terms of an attitude to dementia scale. [33] They noted, however, that these adaptations were likely to have led to under or over-estimation of the relationships between the two variables in their analysis.

2.5.2.2 Baselines and ceiling or floor effects

Identification of change was sometimes problematic because of difficulties around baseline measurement. Authors connected these to issues for recruitment and sampling, and also to the progression of dementia symptoms. The result is indications that activities might not be reaching those who will most benefit from them; a number of authors of included studies reported problems with measurement linked to ceiling or floor effects (the inability of a scale to reflect the upper or lower ends of participants' spectrum of responses to an intervention).

As has been previously noted, in many arts and dementia studies reviewed, there was no clinical diagnosis of a dementia and sometimes only a broadly indicative measure of the level of symptom severity. Sometimes this was presented as deliberate, the result of a decision to focus on inclusion rather than incapacity [48] or because distinctions between different dementias were not thought to influence the particular variables under consideration, such as quality of care. [67] It was also suggested that while a lack of focus on a particular type or level of dementia in recruiting for a study may limit opportunities to draw conclusions about who might benefit from it in future, it could potentially also have the effect of rendering the study's sample more generally representative of the wider population of people with dementia. [50] Further, some proponents of the arts might agree with the authors of a Spanish study exploring the effects of an artistic educational programme, that a diagnosis of dementia should not differentiate participants in a program of artistic activities and living independently in the community from any other adults taking part in similar activity. [57]

However, authors also pointed out that a lack of knowledge about the symptom severity of participants meant that decisions about the kinds of outcomes considered desirable in the first place could not be fully and accurately informed. [46] It was suggested, for example, that it made it difficult to assess the relationship between symptom severity and wellbeing improvements resulting from an activity. [72] The result may be insufficient understanding of how a clinically determined level of impairment or type of dementia relates to an individual's ability to participate in arts activity or the likelihood that they will benefit from it. Additionally, as dementia symptoms vary widely between individuals and for an individual over a given time period, there will continue to be dynamic and subjective constructs, such as resilience, that evaluators will struggle to measure. [79]

Arts projects seeking to improve wellbeing or quality of life may attract motivated individuals whose existing wellbeing, resilience, and access to services is already good, meaning that changes resulting from an activity may be small. A study of a singing group involving people with dementia and their family carers explored, amongst other things, the group's impact on carer quality of life. [8] A failure to show a significant difference against this outcome may have resulted from the study's (self-selecting) recruitment of those who were already coping well. Similarly, caregivers taking part in a chorus for people with dementia and family members, were reported to have had very low levels of some of the outcomes of interest, including depression, at baseline. [77] People with dementia in two care facilities in Australia took part in a trial involving a music intervention designed to decrease depression, anxiety and agitation and to improve quality of life. In the two included studies reporting the findings of this, participants had consistently low levels of all of the negatively measured key constructs at baseline, and good quality of life, with all remaining relatively stable throughout the six-month study. [11, 12]

Dementia is a condition whose symptoms are progressive. As well as noting that this resulted in attrition among participant groups, authors also recognised that, in individuals taking part in an ongoing programme of activities, this progression should be accounted for in attempts to measure change, and that it might compromise the ability of scales to adequately measure constructs such as quality of life over the course of a study. [4, 8, 33]

Because the majority of people living with dementia do not choose or are not able to access services such as those provided by artists, studies will – of necessity and for multiple reasons – involve individuals who do have access to them, and this may therefore also seriously skew study findings (Victor *et al.*, 2016).

2.5.2.3 Measurement of key constructs in the context of dementia

The main outcomes being measured in studies in this review were: those associated with behavioural and psychological symptoms of dementia, including anxiety, agitation, depression and aggression; quality of life (for persons with dementia and care partners); subjective wellbeing (for persons with dementia and care partners, and; impacts for care practices, on professional caregivers and family carers, and carer 'burden' in particular. These outcomes are familiar from research around psychosocial interventions in general. Arts and dementia evaluative studies frequently repurpose measures from studies of such interventions. Studies also attempted to measure engagement or the quality of engagement, either directly or through proxy measures. [6, 11, 12, 15, 21, 29, 34, 57, 72, 75]

Some authors questioned the way in which existing scales operationalised the measurement of particular constructs. For example, the authors of a study evaluating the effects of a programme of elder-clown practice suggested that some of the more complex neuropsychiatric symptoms being explored might elicit nuanced responses from observing raters, while others – such as aggression – might have more concrete indicators that were less problematic to rate. [32] This, it was felt, might contribute to the failure of a study to detect change in the more complex constructs and lead to under or over-estimation of change in those constructs whose indicators are less nuanced or easier to observe. It would also cause difficulties in situations in which an interaction between these or other constructs might need to be considered. These authors also suggested that measures they had used did not reflect current understanding of dementia symptoms, for example, the importance of differentiating between needs-driven behaviours and those which might have other causes, such as pain.

It may also be the case that a global construct, such as wellbeing, measured as a whole, might not show significant change, whereas its subdomains, when analysed separately, will highlight important trends and patterns. [83]

2.5.2.4 Measurement of change in the context of arts activity

The effects of participation in arts activity can sometimes appear to resist measurement. Rather than attempting to measure an outcome such as wellbeing, authors also attempted to value the effects of arts activities using approaches that referred more directly to aesthetic factors. A limited number of tools are available to do this. One (the Creative Expressive Abilities Assessment tool) was used in a study evaluating a programme of musical reminiscence activities. [69] However, the authors found it was not satisfactorily able to reflect an individual's changing patterns of engagement and interaction during an activity session or to account for important contextual elements such as current mood, session theme, musical affinity, the number of facilitators, or the level of interaction involved. Another measure (the Arts Observational scale) was applied in recording the impact of participatory music activities on patients with dementia in a hospital setting. [67] This recorded change according to a number of criteria, including mood, distraction, relaxation, engagement and agitation. However, the authors noted that the tool, while efficient in many ways, still only represented a limited way in which to capture 'tentative impacts'.

The Greater Cincinnati Chapter Well-Being Observation Tool (GCCWOT), designed to assess the psychological construct of wellbeing in the context of arts activity, was developed and tested in one included study [31] and applied, sometimes in adapted form in others. [20, 25, 51, 84] Its adaptations reflect the difficulty of employing a tool devised and tested for one kind of arts activity in other quite different contexts. For example, the authors of one study noted their lack of confidence in the tool's ability to assess wellbeing, given the highly structured nature of the activity they were evaluating. [20] In another study, it was suggested that the tool might simply be measuring ‘well-being’ and ‘ill-being’, with little nuance. [15]

Frustration was also sometimes expressed at the difficulty of capturing subtle aspects of the creative engagement process using quantitative measures. These include those effects that are said to vary according to the social and interactive parameters of an activity, [80] and the sometimes small but important non-verbal exchanges that occur between participants and others during an activity. [20, 33, 56, 66, 69]

It is a common and connected observation that the effects of arts activities for people with dementia may be transitory in nature. As one carer commented of involvement in an art-gallery based intervention detailed in one study, even though the intervention might not reverse the symptoms of dementia, ‘you do it for the moment’. [34]

There are indications that many authors are attempting to capture transitory or in the moment effects. Those that did not often recommended methods that would enable this in future studies. Researchers and evaluators cited the need to address the question of what happens during engagement in arts activity as the main reason for their use of observational methods. As was noted in section 2.4.3.2 above, almost half of all included studies used a measure or tool designed to capture what was going on during an intervention, as opposed to measuring change.

2.5.3 Theme 3: People and settings

Perhaps the most pressing practical evaluation challenge unrelated to study design, and reported within the literature, is that of how data can be collected reliably from people living with dementia.

Findings related to this theme and its sub-themes are listed and described in Table 4 and reported below.

Table 4: Theme 3: People and settings

Theme	Sub-themes	Description
People and settings	Accessing first-hand accounts of people with dementia Caregiver or proxy reports Ethical issues Activity settings, organisational culture and routines	The nature of the symptoms and the condition of dementia created barriers and limited the involvement of people with dementia in evaluation. Alternative methods, such as a reliance on caregiver or proxy reports were problematic. Ethical issues were reported that may be leading to a lack of representation. Activities are taking place in environments that are complex and demanding, requiring evaluators to be attentive to organisational culture and routines.

2.5.3.1 Accessing first-hand accounts of people with dementia

Communication with participants with dementia can be challenging for evaluators and researchers. Study authors demonstrated widespread agreement about the difficulties of using self-report measures or requiring assessments of subjective constructs with participants whose dementia had reached moderate/severe levels of cognitive impairment. The authors of one study questioned whether people with dementia were able to grasp the varying nature of the concept of quality of life, suggesting that their restricted range of observed circumstances might result in them giving similarly restricted observations of their own subjective condition. [20]

Although it is notable that thirty-six of the included studies did use methods enabling participants with dementia to give accounts in their own words, this means that the majority of studies did not. Authors occasionally reported direct quality of life assessment to be impossible beyond a certain degree of impairment using standard, language-based self-report measures or interview. [74, 79, 80] Participants sometimes also had physical impairments or other health problems, making interviews difficult or preventing them from keeping appointments. [6, 57]

Multiple or lengthy survey instruments, when used, could be reported as particularly problematic. Forms took longer to complete than expected, [5, 20] could be fatiguing or simply too difficult for participants (leading to missing data), [6] or required the support of a helper. [20] Notably however, in a few studies simple situational pictographic [33, 80] or visual analogue scales [71] were described as useful, even at advanced stages of the condition.

Evaluators and researchers noted attempts to lessen burden and fatigue through the application of measures at limited time points. [13, 46] Thirty-two studies employed observation rather than direct interview, with some noting the decreased burden upon participants. The use of video or audio for observation in twenty-three studies may have lessened the burden further, leaving participants largely unaware that they were being observed or were the subjects of evaluation or analysis. However, while direct observation countered some of the problems noted earlier, it was also reported that the limitations of structured observation schedules made it difficult for evaluators to assess impacts outside the observed context or times. Observational methods were also noted to be time-consuming, resource-intensive, and heavily dependent for their quality and reliability upon the use of trained observers. [73, 74]

Semi-structured interview has been recommended as a suitable method for eliciting data from people with dementia in the early to moderate stages. The authors of one included study of an art-gallery based intervention noted that participants had expressed a preference for talking about their experience with someone, rather than responding through a questionnaire. [7] However, other authors reported unsuccessful attempts to interview participants with varying degrees of cognitive impairment, particularly those whose dementia was moderate or advanced and those living in residential care settings. For example, authors found participants to be non-verbal, giving short or fragmented responses, needing frequent prompting, or struggling and sometimes failing to remember the activity at all. [4, 22, 28, 34, 37, 41, 53, 67, 69]

A lack of recall was a particular issue if the interview took place at any distance in time from the activity itself, but sometimes even presented difficulties immediately afterwards. Two studies described problems in eliciting useful data using retrospective telephone interviews. [4, 53] In one exploration of the effects of a stand-up comedy programme with participants with mild-moderate dementia for example, it was noted that while participants were perfectly able to engage in ‘icebreaker' conversation on the phone, none could recall any details of the workshop and performance in which they had been involved a week previously. [53]

2.5.3.2 Caregiver or proxy reports

Some studies noted a need to account for differences between responses given by people with dementia, and those given by caregivers or others asked to act as 'proxies' or to answer 'as if they were' the person concerned. In some cases, while persons with dementia did not report changes in quality of life or wellbeing as the result of arts activity, caregivers reported positive changes. [9, 14, 20, 80, 84] It was suggested that these differences may have related to a tendency in individuals with dementia to 'flatten' responses or a desire to normalise them, particularly when being asked to give them repeatedly, a positive association between decreased caregiver burden (measured and found to have decreased significantly after this particular intervention) and the way in which family caregivers rate the quality of life of those for whom they care. [20] Alternatively, it was hypothesised by the authors of a study evaluating a creative expression programme in China that the differences may have reflected caregivers' need to grasp for even slight improvements positively. [9] It is worth noting that authors also reported that professional caregivers were likely to have had limited time in which to contribute to research or evaluation, leading to difficulties when they were expected to provide ratings for multiple persons with dementia in their care. [13]

Where persons with dementia were interviewed alongside a care partner, the 'dyad' perspective could be unique and useful, [69] and the presence of a care partner helpful in putting participants at ease. [41] However, concerns were also expressed about family carer views being over-represented within studies where they were more cognitively or verbally able than their cared-for partner, leading to them dominating interview discussions. [e.g. 41] It was also noted that caregivers and participants sometimes had quite different understandings of the mechanisms behind observed effects. For example, the observed raised mood of participants in one study was attributed by those participants with dementia to the 'atmosphere' resulting from the activity itself, but by caregivers to the introduction of social contacts. [33]

One author also reported concerns about the measurement of multiple outcomes in studies in which measures were completed by different kinds of respondents (for example, agitation being rated only by carers, anxiety assessed only through self-report), meaning that they might not be commensurable. [11, 12]

2.5.3.3 Ethical issues

Formal processes around ethics review and consent were reported as negatively affecting recruitment of some participants, although this was perhaps not as common as some of the other limitations. In one study, the authors noted insufficient time to apply for ethical approval to involve people with dementia; [66] in another the authors described longwinded ethics paperwork at a first session as leading some potential participants to simply opt out of the study. [76] The authors of other studies described the negative effects of commissioning evaluation research at short notice, particularly in terms of recruitment and ethical processes, including implications for the participation in studies of participants who lacked capacity. [15, 37, 66, 76]

It could be challenging and time-consuming to find appropriate venues and times in which to conduct interviews and focus groups where participants with dementia could feel comfortable and uninhibited in their responses. For example, researchers in an acute setting described participants declining interviews out of anxiety, because they didn't want to risk missing an opportunity to see a doctor, or because they were expecting visitors. [67] Attempts to ensure the evaluation process was convenient for participants led one researcher to diverge from a protocol, conducting more focus groups and individual interviews than planned, and including family members in them. [37]

Researchers and evaluators may have unhelpfully normative expectations for interviews with people with dementia. Commenting on this, the author of one ethnographic study exploring individual experience of people with dementia taking part in a singing group activity, recommended that interviewers should relate and communicate 'proactively' with people with dementia rather than expect to 'listen'. [27] She successfully used an arts-based elicitation technique in interviews, encouraging participants to sing songs familiar to them from the singing group in which they had all participated together.

Evaluators and researchers faced issues with balancing the subjective nature of some of the constructs of interest (such as wellbeing, quality of life, or resilience) with the use of tools or methods describing or measuring their expression in participants who were not able to reflect on their situation or to communicate because of cognitive impairment. A user-led approach for addressing the issue of how to successfully explore the outcomes that are meaningful for people with dementia, was recommended in one study. [36] In another study reporting on development and piloting of a scale to be used to assess the use of music in dementia contexts, it was suggested to be imperative to consider the perspective of the person with dementia and his or her perception of what was important about an intervention, rather than the needs of carers to manage or alleviate behavioural and psychological symptoms. [37] Addressing such concerns, qualitative studies often included aims such as a desire to improve understanding of the wider role that the arts, and their implementation in a programme of activity, might play in the lives of people with dementia. [e.g. 4, 8, 27]

2.5.3.4 Complex, demanding and challenging environments

The setting in which an arts activity is delivered is sometimes directly described as a mediator for, or contributor to its effects. [8, 23] There were six studies including an explicit focus on the process or implementation of arts activity. [1, 16, 63, 64, 70, 78] Two of these concerned the same ‘living room theatre performance’ activity, exploring it as an intervention for people with dementia in residential care [16] and as a contact method used by care staff. [63] The others examined: potential barriers and facilitators to implementation of a personalised music activity delivered by home professional caregivers; [1] factors affecting facilitation of an arts programme in residential care from the perspectives of the artists involved; [64] the implementation of an interactive art programme in twelve Dutch museums, and; [70] the design and implementation of an art-making programme for people with dementia and their caregivers living at home. [78]

The dynamics and complexities of a setting can function to enable or restrict the evaluator. In several studies, a lack of definition in relation to how much of an activity was delivered or to the length of sessions (sometimes referred to by authors as a ‘dose’), was described as a limitation. [4, 32] Working with existing programmes can lend the results of an evaluation ecological validity; in one study the authors reported making a deliberate choice to replicate what they termed a standard ‘dose’ outside the research context of just one session of group singing activity per week, rather than evaluating something that would otherwise prove impractical for others to deliver. [13]

Otherwise promising findings were sometimes described as having limited transferability to other contexts because the evaluated intervention had been implemented with a small sample in a single site or in a limited selection of settings. Unsurprisingly, evaluation was found to have been difficult when a project involved multiple and contextually different sites and groups receiving varied iterations of activity. [84] Evaluators faced with the challenge of trying to design a study whose results might be applied across different settings sometimes attempted to recruit across multiple sites. Some authors noted that this risked the introduction of further difficulties, including the loss of key contextual information if participant data were aggregated across sites.

Only four studies involved activities taking place in either acute [3, 67, 84] or day hospital settings. [37] The recruitment of suitably sized and stable groups involving people with dementia in these environments can be particularly challenging because of the fluidity of the hospital population and the demanding routines of staff. [84] This led the authors of one study to investigate alternative research designs and recruitment options, including the analysis of routinely-collected ward-level data to complement observational data and interviews and focus groups with individual patients, visitors, the musician, and hospital staff. [67]

Nursing and residential care settings made specific and multiple demands on evaluators and researchers. Where participation over the course of an evaluated programme was not stable within a care setting, various contributory factors were reported. These included challenges resulting from implementation of the arts activity and its interaction with the setting’s systems and routines. Evaluators were not necessarily able to control elements occurring around the arts activity.

Potentially contextually significant factors were consistently under-explored in the literature, for example, the care philosophy that might have been prevalent in a particular care setting. However, a sensitivity to the expectations and requirements of the culture and routines of the activity setting was seen as playing a role in decisions about the methods and conduct of studies and recognised as a contributory factor in their success or failure. For example, the timing of activity sessions might be either barrier or facilitator, particularly if residents needed support from care staff to attend. In a pilot randomised controlled trial in Australia investigating the feasibility and effects of a dance program for people with severe to moderate dementia in a nursing home, dance was compared to music appreciation and socialisation groups in the same setting. [75] It was found that timing and the home's routine may have played a role in determining both the intervention's feasibility and its effects: the comparison activity was delivered immediately after breakfast, when most participants were already in place and alert, whereas the dance activity was scheduled late morning when residents were often back in their own rooms and some were asleep, meaning that they had to be woken up and escorted to the session. In one study in a day care setting, it had been planned to observe participants engaged in another more traditional activity as a form of comparison; however, the content of the comparison activity was not determined by the researchers, and sometimes it was replaced by an outing, making its observation impossible. [31] In two studies, infectious outbreaks were said to have hampered data collection. [32, 35] Activity sessions were also reported as being scheduled around care home routines, [13] or interrupted by building work or for personal care. [15]

Outside residential care contexts, the procedural dynamics of a setting also sometimes played a role in determining the methods that could be applied. When activities were delivered in public, for example, in an art gallery or museum, evaluators sometimes found themselves limited to particular kinds of observation, including being unable to use video because of ethical constraints around filming in public spaces. [60] In another art museum setting, procedural considerations led the evaluating team to recommend that future studies used time series analysis methods alongside a longitudinal design to overcome challenges they had faced in understanding the interaction between the setting and the intervention. [80]

2.6 Discussion and further synthesis: The warp and weft

The findings above demonstrate that the limitations and challenges reported by study authors are often framed by the requirements of their chosen research paradigm, although it is evident that authors may also ‘borrow’ criteria (such as generalisability or bias) from other paradigms. The next step for this review was to identify whether there might be themes and concepts relating to methodological challenge operating for evaluators and researchers within the published studies, regardless of research paradigm or methodology, and also perhaps regardless of whether they were openly reported. To explore this, I returned to the reviewed literature for a closer systematic reading and synthesis, as described above (section 2.3.3).

As a result, I identified five sensitising concepts relating to methodological challenge and operating across and between the included studies. These are: Value, Context, Ethics, Meaning, and Use. I suggest the metaphor of *warp* and *weft* to describe how these relate to the areas of challenge described above. In weaving, the warp is the tensioned thread running vertically through the cloth. It is there from the start, and its characteristics and nature (colour, weight, texture, tension, distance apart, and so on) are set up on the loom. The weft is woven through these warp threads; the way in which it does this, as well as the qualities and nature of the weft thread, will differentiate one piece of fabric from another.

The *warp* threads are, in my metaphor, the areas of tension and challenge running through any evaluation or evaluation research project – my sensitising concepts. They will be rendered more or less visible as a result of the technical applications of the methods woven through them – those factors relating to study design, the measurement of change or understanding of experience, or the challenges of involving people with dementia or working in arts and dementia activity settings. To return to my typological discussion earlier in this chapter (section 2.4.3.1), they will also differ according to the kinds of research question being asked. The resulting *fabric* of methodological challenge is different for each study, although there are likely to be similarities and consistent patterns.

I present and describe these important threads below and refer back to them throughout the remainder of this thesis.

- **Value:** Issues relating to differences in value and value judgements – both in terms of the kinds of evidence created through evaluation, and the value of arts and culture itself.
- **Context:** Differences of understanding around the role played by context in evaluation.
- **Ethics:** Concerns relating to the responsibility and duties of an evaluator, the rights of participants, and their expression within formal ethical processes.
- **Meaning:** Problems in understanding what might be meaningful about engaging in the arts for people with dementia.
- **Use:** Issues relating to the purpose of evaluation, and the way in which the knowledge gained is used.

2.6.1 Value

From attention to the aims and outcomes in the published literature included in this review, it might be concluded that arts activities are valued variously because they: provide an alternative to pharmacological interventions aimed at ‘management’ of the behavioural and psychological symptoms of dementia; can deliver specific and measurable cognitive, social or emotional benefits for people with dementia; can improve communication and interaction between professional or family carers and people with dementia or lessen the ‘burden’ of care; can enhance the wellbeing or quality of life of people living with dementia and/or their family caregivers; or, offer opportunity for engagement in activity that is enjoyable, personally meaningful or aesthetically pleasing for individuals. Economic justifications for the value of arts activity were not explored in any of the included studies.

Differing conceptions of instrumental or intrinsic value of the arts, and the potential role of arts and creative practices within the lives of people with dementia, are reflected in researcher’s choices of study design and methodological approach. However, value – of art or of the virtues of a chosen paradigmatic approach for research or evaluation – was rarely discussed explicitly. The reasons for choosing one conception of what is valuable over another were rarely questioned or otherwise explored. This suggests a gap in the reported literature whose implications would benefit from further unpacking. This lack of exploration may be because of the widespread acceptance of particular hierarchies of evidence, or because what we choose to research is ordered in terms of 'the cultural values with which we approach reality' (Weber, 1949, p. 78). It is perhaps the case that a lack of acknowledgement of the role of value among some study authors is a facet of their chosen research paradigm, and this may be having an effect on their conduct of evaluation or research.

2.6.2 Context

For the purposes of evaluation research, context may be defined as the diverse set of circumstances or factors that surround an intervention or its implementation (Damschroder *et al.*, 2009). How an evaluator or researcher will deal with context and detail is an important methodological decision. Study authors highlighted multiple variables deserving consideration and affecting the reliability of results from experimental and non-experimental studies. The prevailing paradigms in the field have often led to such details being viewed as variables to be controlled and accounted for.

An alternative view is that: context and confounders may 'lie at the very heart of the diffusion, dissemination, and implementation of complex innovations. They are not extraneous to the object of study; they are an integral part of it' (Greenhalgh *et al.*, 2004, p. 615). In addition, lack of clear description of interventions has been described as a weakness of evaluation and research studies in the field, potentially obscuring our understanding of the mechanisms at work (de Medeiros and Basting, 2014). Attention paid to context and detail might facilitate better understanding in this area. This review demonstrated that, whether implementing an activity or evaluating it, documenting or capturing appropriate contextual information and accounting for its effects remains difficult. A consideration of the reasons for this, and attention towards ways in which to address them, will be useful in identifying potential solutions to methodological challenges.

2.6.3 Ethics

Study authors engaged with questions about how consent might be informed, consideration of what did and did not constitute capacity for a person living with dementia, and the repercussions of these questions for the methods they employed. However, this review suggests that the ethical concerns of evaluators and researchers encompass not only the common practicalities of consent and ethical review processes, but that they are also increasingly extending into questions around how evaluation research can meaningfully reflect and empower the varied voices and perspectives of people with dementia.

While many included studies did include accounts of their first-hand experience, many of these were limited, and people living with dementia do not appear to be routinely or meaningfully consulted about the development, implementation, or evaluation of arts activities described in the studies in this review. Therefore, whether and how to engage them as active participants in these endeavours will constitute a particular area of interest for future study.

2.6.4 Meaning

As we have seen in the reviewed studies, evaluators and researchers reported significant difficulties when applying all the methods conventionally used to access the experience of subjects with people with dementia, including interviews, focus groups, surveys and questionnaires. Decoding what Husserl termed ‘the enigma of subjectivity’ (Husserl, 1970) is one of the most fundamentally challenging aspects of any research attempting to understand the psychological responses of human subjects. Subjectivity cuts both ways – this review showed that evaluators and researchers faced challenges in considering both the effects for their work of their own subjectivity, as well as that of their participants. This is particularly pertinent because communication between the researcher and the subject of research is likely to be further compromised as a result of the limitations to cognitive ability and language skills, lack of orientation to place and difficulties with attention that can affect individuals living with dementia (van Baalen *et al.*, 2010). Understandings relating to meaning – what is meaningful about participants' engagement with art and their responses to it, and what is meaningful to evaluate – particularly affect decisions made about approaches, methods and tools that a researcher or evaluator uses, as well as the kinds of relationships they can have with the subjects of their research.

2.6.5 Use

This review's final connecting concept concerns the translation of evidence into practice or how the knowledge gained through evaluation is used. The authors of published studies sometimes referenced the strategic or policy context for activity when stating their aims and objectives, and made recommendations relating to future practice and research. But many of the key questions highlighted by the typology of evidence referenced above were not addressed through the included studies, suggesting a gap in terms of evaluation research designed to address questions of knowledge use.

The full implications of what it might mean to be conducting evaluation rather than research only became clear through the grey literature 'sense check'. Much of this grey literature was commissioned by arts organisations, and therefore it included material that was more closely attentive to issues around arts practice and its implementation, and to the experiences and needs of artists and arts organisations. There were indications of a desire to share good practice, to disseminate evaluation findings for information, and to inspire and to advocate for the evaluated project and the organisation delivering it. These intentions are reflected in accompanying practical guidelines, frameworks and toolkits produced to guide arts practice, and recommendations included within evaluation reports to support future evaluations. This focus represents a perspective largely missing from the published material – that of the arts organisations and artists involved. Having the translation of knowledge, evidence or research into practice as a motivational factor for evaluation might have an effect on methodological decisions and expectations surrounding it. However, the use of designs that are inappropriate to addressing these aims may be contributing to the methodological difficulties experienced by evaluators.

In Chapter 2, through a review of the literature, I described the range and scope of methodological challenges experienced by evaluators. Methodological difficulties were identified, along with a provisional set of sensitising concepts (value, context, ethics, meaning, and use). I used a weaving metaphor to describe how these concepts relate to each other. The following chapter engages with some of the methodological implications of the review's findings for my research and it describes my theoretical influences. Outlining my chosen methodological approach, I argue its appropriateness for the research question and frame its particular ontological and epistemological understandings. This methodological discussion is applied in Chapter 4 within a description of the methods, design, ethical concerns, and other considerations for the study.

3 METHODOLOGY

3.1 Research question

The subject of this thesis is the methodological challenges involved in the evaluation of arts-based activities for people living with dementia. My research question is:

What is it about arts-based activities for dementia that causes evaluators and researchers problems when describing, explaining, justifying, and implementing the methods they use to evaluate them?

This question reflects both a philosophical approach and a particular definition of methodology. The approach and the reasoning behind the question is described in detail later in this chapter. The particular understanding of methodology implied was outlined earlier as part of a definition of the term applicable both here and throughout the thesis as a whole (section 1.4.1).

3.2 Aims and objectives

This study aims to examine and respond to calls for greater methodological rigour in arts and dementia evaluation and to identify some potential solutions. It does so through an exploration of challenges to evaluation in practice. The enquiry is qualitative in nature and emphasises the importance of stakeholders' experiences.

Its findings are described and interpreted in the light of an epistemological orientation whose theoretical influences will be outlined in this chapter.

In achieving the study's aim, I take an in-depth look at the nature of methodological challenge within the literature. Gaps and themes exposed are then interpreted and used as sensitising concepts that inform an exploration of the experiences of individual stakeholders involved in the evaluation of arts and dementia activity. A spectrum of views is introduced in order to avoid the biases inherent in looking at the subject from one perspective. The information derived is then discussed and used to signpost solutions.

3.3 Theoretical influences

In the Introductory chapter of this thesis the influence of the evidence-based medicine (EBM) framework upon the development of arts and dementia evaluation, research and practice was noted (section 1.2). The principles of evidence-based practice also represent a key theoretical influence on my own research.

Although the word *framework* has been used above, *paradigm* could also apply, a use of the word in relation to scientific research that was originally coined by the philosopher of science, Thomas Kuhn. He termed a paradigm a theoretical orientation and a reflection of the interests of a knowledge-seeking community at a particular historical moment (Kuhn, 2012). In the particular historical moment in which I conduct this research and for the field about which it concerns itself, the positivist influences that inform evidence-based medicine are currently more dominant than others (Knight et al., 2017). Their methods are likely to be judged as both more accurate and perhaps more influential in policy terms than the alternatives (Jerrim and De Vries, 2017). An instrumentalised, outcomes-focused approach continues to speak convincingly to the needs of policy-makers and those who fund and commission arts and leisure activities for health and wellbeing within the public sector.

However, while EBM's well-established and tested processes and frameworks for evaluation and research are widely accepted within health and public health contexts (Craig *et al.*, 2008), there are those who suggest that our interpretation of it in practice should be more flexible (Greenhalgh, Howick and Maskrey, 2014). Others highlight the viewpoint that clinical decision-making, as opposed to public policy at least, is ‘relentlessly situated and contextual’, recommending a perspective for decision-making that starts from the individual patient, rather than the abstracted results of research (Wieringa *et al.*, 2017).

Of course, other paradigmatic research orientations are available to those working in the field of arts and dementia. These include participatory approaches that emphasise ideas of co-creation and ‘human flourishing’; approaches that emphasise the phenomenological ‘lived experience’ of people affected by dementia, and; critical approaches that focus on tackling inequalities resulting from, for example, disability or malignant social psychology, and which urge us to pay attention to capacity rather than incapacity, or which aim to uphold human rights and the citizenship of people living with dementia. However, the literature review showed that these are currently less well-developed in terms of their influence on evaluation research in this field.

It is unwise to ignore the impact of the status quo on researchers when trying to understand what is happening in the field. However, following Kuhn (2012), I suggest it is also unwise to assume that what appears normal now will seem so in the future. And, as the interdisciplinary sociologist John Law has suggested:

> […] if we build our assumptions about the nature of good methods into our investigations of method then we are likely to come to conclusions that mirror those assumptions. We are likely to find that ‘good methods’ produce ‘good results’. We will tend to reproduce the current workings of method. (2004, p. 40)

It can be difficult for those who come to arts and dementia from disciplines where alternative paradigms are prominent, such as the arts and humanities, to work within or even alongside the EBM paradigm. As my literature review showed, it is common for evaluation research work in the field, even when sympathetic towards alternative orientations, to be framed in evidence-based and largely positivist terms. As a result, practitioners of all kinds can experience translation difficulties when either commissioning, conducting, or attempting to learn from evaluation, leading to misunderstandings of various kinds (Daykin *et al.*, 2016, 2017). There may be a need for new vocabularies to describe the way in which method is to be enacted in the field.

This seems a good point at which introduce two further theoretical influences informing my perspectives in this research, both of which hail from the field of Science and Technology Studies. One is exemplified in the work of philosophers and sociologists Bruno Latour and Steve Woolgar, in particular their description of the social construction of science through an anthropological observation of it in routine practice (1979). The other is the delineation by John Law, already introduced above, of the concepts of 'method assemblages' and the 'hinterland of method' (2004).

In exploring 'laboratory life' Latour and Woolgar steadfastly refused to take scientists' words about their work as gospel. Instead they observed the activities of working scientists in a particular laboratory as if they were a little known tribe, examining the way in which labour was structured and divisioned, the 'arrangements' between people and things, the way in which the eventual products of laboratory life – its written papers and reports – come into existence, and what these written 'traces' revealed about the process of their production, or the 'construction of a fact'. They discuss the subject of scientific motivation, advancing a description of the concept of *credibility* through its theorisation within an endless investment cycle of credit. This cycle is described as making possible a conversion between 'money, data, prestige, credentials, problem areas, argument, papers, and so on' (1979, p. 200), accounting for the currency of scientific fact in both economic and epistemological terms.

Law further develops the insights of Latour and Woolgar in discussing how knowledge is produced. Method, as he describes it, extends endlessly, beyond the limits that we usually imagine for it. His method 'hinterland' is evident in the bundles of social and material relations associated with knowledge production, which he terms 'assemblages'. These incorporate and extend beyond 'laboratory benches, reagents and experimental animals, or questionnaires, interview design protocols, and statistical or qualitative data-analysis packages […] into tacit knowledge, computer software, language skills, management capacities, transport and communication systems, salary scales, flows of finance, the priorities of funding bodies, and overtly political and economic agendas' (2004, pp. 40–41).

For Law, a method ‘assemblage’ includes the relations between elements traditionally understood as ‘gathered together’ by scientific work, and this ‘Othered’ set of elements which are not conventionally understood as forming part of scientific method.

These theoretical influences suggested to me the need to think hard about what an evidence base is, what it is intended to do, where its limits extend, how its facts are created, and what the motivations might be for evaluators and researchers who work towards supporting and extending it. I felt the need to be aware of how the discourse of evidence and its surrounding paradigmatic apparatus might be affecting the way in which evaluation research is conducted in the field of arts and dementia, the kind of evidence that is being produced, and the way in which it is taken up by those who wish to use it in delivering arts and dementia practice, to benefit people with dementia, or to influence policy.

To understand how stakeholders in evaluation (artist practitioners, evaluators and researchers, arts managers, those who fund, commission or make policy, and participants themselves) can best address the methodological challenges the field faces currently and in the future, I argue that there is a need to explore the figurative and material landscapes that these stakeholders inhabit and which, if we agree with Law, their inquiries also create.

This kind of exploration can be achieved through reading the texts and other materials that they produce, by observing, and also by asking them directly to think and talk about, their experience, their practice and the tools they use to navigate, as well as the barriers and difficulties they experience. Analysis of these data could then usefully be focused on understanding how methodological challenge comes to be manifested in the way that it is for particular individuals in particular contexts as well as more generally in the field. My hope is that this type of inquiry will be more fruitful in identifying solutions that originate within arts and dementia evaluation practice, than would be a quest that sought to determine the quality of the methods evaluation researchers currently use, or which attempted to fit them more closely into the framework of a particular paradigm or discipline.

3.4 Identifying a philosophical framework

The process by which these conclusions about the aims and objectives of the study have been reached, and through which a methodology was chosen to support them, was not entirely straightforward. Prior to the literature review, the study's proposed research questions were: (1) *What are the methodological challenges facing evaluators of arts-based activities?* and; (2) *How might these challenges be addressed?* As was described in Chapter 2, the question that eventually drove synthesis of the texts in my literature review was: *What can we understand from a review of evaluation and research studies about the challenges facing evaluators of arts-based activities for people with dementia?*

The literature review revealed – as expected – a multiplicity of practical and technical challenges for application of method and measurement. Some of these are generic to particular kinds of study design but may be particularly prevalent in this field. Some can be more directly related to the way in which local contexts (the particular arts activity, delivered by a particular artist, involving particular participants, taking place in a particular setting) impact on and might affect evaluation design.

The review also provided me with the metaphor of the *fabric* of methodological challenge, in which five sensitising concepts form the tensioned warp threads, through which the weft of method choices weave. Some of these tensioned threads were specifically referenced or alluded to in reported study limitations, but some were only visible through absence of specific discussion about them in individual studies, despite their being suggested as concerns elsewhere, or within wider discussions within the field of arts and health.

To reiterate, the sensitising threads, or concepts are: issues relating to the effect of differences in concepts of value and value judgements might affect methodological work; differences of understanding around the role of context; concerns resulting from, and navigation of, the evolving ethical landscape around the condition of dementia and the involvement of people with dementia in research and evaluation; problems in understanding what might be meaningful about engaging in the arts for people with dementia, and; issues relating to the purpose of evaluation, and the way in which knowledge gained through it might be used.

The end goal of the research has been from the start to contribute towards development of effective evaluative practice for the field. When first considering what research approach to take, and following Flyvbjerg (2011), I saw concrete, context-dependent knowledge as key to such an endeavour. I felt that some of the most intractable methodological challenges for the field might be directly related to the contextual variability of project implementations and the difficulties researchers find in making sense of the inherent flexibility of the modes of arts delivery, the multiplicity of different outcomes that have been associated with the arts and the necessity of capturing impacts in the moment. Therefore, when initially proposed, my intention in the study was to explore the questions using a mixed methods *phronetic* case study approach, involving ethnographic observation of projects and barriers and facilitators to their evaluation in practice. This would have created rich data for analysis and could have been followed up with in-depth interviews with project stakeholders.

This approach was certainly an appealingly interesting prospect for a researcher, and had much to recommend it, methodologically. However, following the literature review, I decided that it would prove limiting. This was partly because of its focus on practical challenges for the application of method, many of which the literature review demonstrated were already well recognised. It was also because of logistical difficulties posed by the need to select appropriate cases while time and resources to identify and then research within these case sites would be limited. Insights from the literature review discussed above, including the sense that there were likely to be significant gaps in the reporting of methodological challenges experienced by researchers and evaluators, suggested to me that there also might be aspects of arts and dementia practice and its evaluation research hinterland limiting the types of methodological problems that were visible.

Recent research has explored the pressing problems of how we might wish to develop more sensitive measurement tools to understand and capture impacts and benefits for people living with dementia within arts and creative contexts (Johnson *et al.*, 2017; Dowlen *et al.*, 2018). Whilst conducting this study, I was aware that research was also underway into understanding how to identify and differentiate between the particular processes and characteristics evident in different art modalities, with the understanding that this would be vital in ensuring that we better capture the actual mechanisms of an individual arts activity in order to assess its outcomes (Cousins *et al.*, 2019b). Such work, although of great value, could not address questions of how to overcome the challenges being reported by evaluators and researchers without accompanying exploration of contextual, systemic or structural barriers likely to affect its application or reception and prevent the uptake of evidence derived from its use.

To lay a useful foundation for future research and evaluation practice, I felt my study needed to encompass three objectives. Firstly, it should enable conceptualisation at a more abstract level of the various challenges experienced by different stakeholders for evaluation. Secondly, it should allow a drawing out of the connecting relationships between these concepts, the structures through which arts and dementia practice is commissioned, delivered, and experienced and what drives the methodological decision-making processes engaged in by evaluators and others. Finally, it was likely to be important to understand how different disciplinary and sectoral perspectives and assumptions might be affecting the fabric of methodological challenge.

3.4.1 Critical realism and realist social theory

It is often considered useful to say that there are a number of different paradigms, basic sets of beliefs that guide action, into which research approaches may be fitted. Over the years these paradigms have been variously categorised and named. Two examples of such categorisations include: *positivism*, *post-positivism*, *critical theory*, *constructivism*, and (in later revisions) *participatory* (Denzin and Lincoln, 2018), and; *predict*, *understand*, *emancipate* and *deconstruct* – this latter from a feminist perspective (Lather, 1992). Underneath the differing terminology lie similarities, but the nuances that divide them are always important. Paradigmatic controversies, questions about exactly how a set of beliefs informing particular kinds of enquiry is constituted, named and defined, and the role that will be played by differing perspectives and disciplines in shaping how they are characterised in use, form part of a continuing and sometimes fiery debate about the landscape of research.

One approach, critical realism, has been described as occupying a middle ground somewhere between the poles of hard science or *naïve* positivism and sceptical postmodern constructivism (Erickson, 2017). It is a meta-theory, meaning that it has particular things to say about the nature of reality (ontology), and also about how we gain knowledge about reality (epistemology). Specifically, it is ontologically realist but epistemologically constructivist or relativist. A critical realist understands the world as being composed of sure and certain things while recognising that any knowledge of those things is constructed by the person doing the 'knowing' and is therefore shaped by their perspective or world-view. This is also broadly the position that I take, and the one which underpins this research.

The figure most closely associated with critical realism is the British philosopher Roy Bhaskar, whose ideas were first set out in his 1978 work, *A Realist Theory of Science*. As a philosophy, critical realism's origins lie in a critique of positivism, the theory that the world has a singular reality about which sensory experience is the only source of certain knowledge, and whose epistemological hallmark is application of the scientific method. One of critical realism's main arguments lies with what Bhaskar termed positivism's 'epistemic fallacy', or what he saw as positivism's attempt to argue that it is possible to reduce statements concerning the world to those about our knowledge of the world. The philosophy has evolved considerably, through Bhaskar's own later writing, and through the work of others, including usefully for this study, the social sciences in the form of realist social theory (e.g. Sayer, 1992, 2000; Archer *et al.*, 1998; Danermark *et al.*, 2002). Realist social theory allows for an understanding of how people (or agents), while they may be conditioned by their structural and cultural contexts, have the potential to instigate change through internal reflexivity and interaction with other agents (Porpora, 1998; Danermark *et al.*, 2002; Archer, 2007).

Critical realism has been critiqued and then used as a significant stepping off point in the development of methods for realist evaluation, synthesis, and review (Pawson and Tilley, 1997; Pawson, 2013). The philosophy comes in a multitude of flavours (for an introduction to many of the most significant ideas, see Archer *et al.*, 1998). In this thesis the term 'critical realism' is used in a broad sense to mean a philosophy whose distinctive feature is a denial that sure and certain knowledge of the world is ever possible, despite the real existence of things and ideas in it (Maxwell, 2012).

3.4.1.1 Reality, causality and criticality

I have identified several further key elements of critical realism as of importance for this study. These are: its description of reality as stratified, differentiated and changeable; its treatment of causality, and; its criticality. I will briefly describe these in turn.

Unlike other postmodern critiques of positivism (such as constructivism or critical theory), critical realism privileges the ontological question *'what must the world be like to make it a possible object of knowledge?'* (Danermark *et al.*, 2002; Bhaskar, 2008, pp. 21–24), rather than the epistemological *'how is knowledge possible?'* This highlights a desire to explore and understand generative mechanisms – the properties that cause events – rather than the observable (empirical) events themselves. It also encompasses the epistemologically relativist position outlined earlier. Bhaskar identified three differentiated ontological domains: e*mpirical* (in which we find the things that we experience), *actual* (in which things continue to happen whether or not we experience them) and the *real* (in which the mechanisms that generate events may be found) (Bhaskar, 2008, pp. 56–62). These mechanisms or properties may be directly observable, or they may not, since they can include mental ideas and beliefs. Importantly, for critical realists, the empirical domain (generally the site of research data) is considered to be always mediated for us through theory. As in other postmodern philosophies, all theories are seen as grounded in a particular worldview or perspective. All knowledge is therefore partial, fallible, and incomplete. Just because we cannot access empirical reality except through theory, however, does not mean that the mechanisms generating events are any less 'real'. This is the case even if a particular theory is shown to be incorrect. While our experience of the world is shaped by theory, it is not dependent on it. Something (the generative mechanism) would still make the apple fall when dropped, even if our understanding of gravity was shown to be wrong.

Whether it is ever possible to assign causality is a vexed philosophical question and one that is of particular importance for empirical research. It has been suggested that in a post-positivist, post-modern world, causality is simply not a viable concept since '[e]verything influences everything else, in the here and now' (Lincoln and Guba, 1985). However, it has also been argued that a realist approach to causality can provide a framework for carrying out and justifying causal qualitative research, and for overcoming criticism levelled against such work from the positivist camp (Maxwell, 2008, 2012). In particular, Maxwell suggests that critical realism enables the researcher to move from a *variance theory* view of causality to one of *process theory*. The variance theory view forms the basis for most experimentation. It treats the process by which inputs and outputs come to differ as unknowable, a 'black box', proposing instead an exploration of the variables themselves and the difference between them, ideally through rigorous and suitably controlled empirical methods. The alternative process theory approach seeks to open up the black box in order to 'establish the qualitative nature of social objects and relations on which causal mechanisms depend' (Sayer, 1992, p. 3).

A method that allows identification of causal mechanisms has clear benefits for research conducted in a complex, messy, hard to control social world. Being able to evaluate successfully under these conditions would avoid, for example, what Pawson and Tilley have called the 'heroic failure' of the experimental paradigm in evaluation – its frequent inability to produce findings that have anything definitive to say about how interventions can be successfully implemented in future (Pawson and Tilley, 1997). For the current study, which is not evaluation, an orientation towards identification of causal mechanisms would still usefully enable me to open up the relationships between methodological challenges experienced by evaluation researchers, the contexts in which they are experienced, and the surrounding theories and other structures affecting them.

In his book explicating (and disagreeing with aspects of) the philosophy of Bhaskar, Collier discusses why it might be necessary to have a 'realist philosophy', rather than to just *be* a realist in the practice of science or, indeed, everyday life (Collier, 1994, pp. 16–20). His answer is that the alternative to philosophy is more likely to be 'bad' or 'unconscious' philosophy than no philosophy at all. To practise philosophy, he says, is to desire to explicate 'knowledge that is already implicit in some practice or other'. It is putting 'criticality' to work in order to illuminate practice. If we take the view that there are social realities, with a structure that is theoretically and historically constituted, and mechanisms whose reality is certain, but whose visibility depends on the epistemological practices and processes used to access them, then the intent of a 'critical philosopher' will be to make visible those structures and their contradictions. It may also be specifically 'emancipatory', meaning that it has a concern with enabling social justice or a desire to change existing practice. While the critical realist doesn't necessarily favour any particular ideological structure, concerns of this kind are likely to be reflected in her research productions, as they are in the work of critical theorists taking a specifically Marxist, feminist, or post-colonialist perspective, for example.

3.4.2 Retroduction

The critical realist's main analytic tool, one which may be used to identify causal mechanisms, is *retroduction*. This has been described as argument 'from a description of some phenomenon to a description of something which produces it or is a condition for it' (Collier, 1994, p. 22). Retroductive analysis involves asking the critical realist question 'what must be true for this to be the case?', abstracting potential causal components, re-describing them in terms of pre-existing potential causal theories ('analogies and metaphors' in Bhaskar's words) and then testing them, somewhat in the way a detective might (Collier, 1994, pp. 122–3). It has been described as a dynamic process of 'iterative abstraction' (Yeung, 1997), one that involves movement of thought, back and forth between observable phenomena and potential explanations for these, in an endeavour to expose the most plausible generative mechanism at work.

3.5 Identifying a methodology

My literature review revealed that the field does, indeed, perceive evaluation research to be methodologically challenging. It helped to identify some broad concepts relating to the nature of these challenges and indicated the existence of particular concepts and contexts that it would be useful to investigate further. It suggested that some challenges experienced were common to evaluation research generally and some to work within the wider arts and health field. It confirmed for me that my research into methodological challenge needed to engage with these general challenges in order to understand what might be particular about the arts as experienced by and with those affected by dementia.

Exploring my own philosophical inclinations in connection with critical realism helped to confirm that the focus of my inquiry, given that its aim was to provide useful navigation aids to other evaluators, should be on the identification of connection(s) between the challenges faced by individual evaluators and researchers, the dementia related contexts in which these challenges are experienced, and the wider structures within which evaluation and arts and dementia practice is commissioned, delivered, or received.

As mentioned above, critical realism is a meta-theory rather than a methodology. While it may be helpful in providing a philosophical framework, and in developing a question, it is not entirely clear how it should be used in the justification and application of research methods (Maxwell, 2012; Fletcher, 2017), outside realist evaluation where the methodology is more developed (Pawson and Tilley, 1997). To begin the process of identifying a research design and methodology that would be underpinned by critical realism, I revisited the original research questions and re-shaped them into a single question that was specifically critical realist in form:

What is it about arts-based activities for dementia that causes evaluators and researchers problems when describing, explaining, justifying, and implementing the methods they use to evaluate them?

I then returned to the key facets of critical realism outlined above: its conceptualisation of the relations between structure and agency, and its accounts of reality, causality, and criticality. I considered these in conjunction with the other theoretical influences already described (the ‘social construction’ of science and ideas of method ‘assemblage’ and ‘hinterland’). The requirement was to establish a methodology that would reflect these and, at the same time, address the study's research question. The following elements played a role in informing my choice of methodology:

1. It would need to be qualitative in approach. It should be capable of exploring the perspectives of those involved in evaluation, the nature and quality of their experience and the meaning they attached to it, and the reality of the social and mental processes that might determine their actions and interactions.
2. It would allow elements of reflexivity, recognising that the conduct of research has real repercussions and is not a neutral act.
3. Its analytic strategy would not be deductive, since this would risk reinforcing normative assumptions.
4. It should allow use of what had been learned from the literature review and prior reading, albeit perhaps as a point of departure for data collection and analysis.
5. It would involve a dynamic analytic process that moved from description and categorisation of any concepts identified to drawing connections between them.
6. It would support the potential for these connections to be of a causal kind. In doing so it would allow commentary on how the nature of arts-based activities delivered for people with dementia, including the contexts and structures that surround them, might relate to the nature of the challenges experienced by those involved in their evaluation.
7. It should aim to expose contradictions, unpick conundrums, and have genuinely useful impact for arts and dementia practice and evaluation.

Ethnographic approaches have been used in critical realist studies (e.g. Lipscomb, 2014; Decoteau, 2016), and would have been appropriate if I had still been considering a case study design. Thematic analysis might have complemented a critical realist approach in the form associated with Braun & Clarke (2006) Although there are concerns over the method's validity, Braun & Clarke argue that these are mainly the result of its misuse and suggest that it offers a robust, systematic framework for coding qualitative data (2014), particularly when accompanied by an appropriate theoretical orientation. Useful as it is for analysis however, its methods do not offer guidance on how to approach the design of a study as a whole and it is not intended as a stand-alone method for the development or testing of theory.

My chosen response has been, instead, to develop a study at whose core is a critical realist approach to the application of grounded theory method.

3.5.1 Grounded theory

Grounded theory originated in the 1960s through the work of Glaser and Strauss. Their aim was to create a research process that would successfully enable the discovery of 'theory from data systematically obtained from social research' (Glaser and Strauss, 1967, p. 2), what Glaser later called 'a total methodological package' (1999). It was developed in response to the dominance of the deductive method (the testing of hypotheses derived from theory through empirical experiment) within the social as well as the natural sciences at the time. Incorporating elements of philosophical pragmatism and symbolic interactionism from its two founders, it emphasises the uncovering of the basic processes at work in a social situation, with rigorous attention to method, and the specific aim of developing substantive theory *grounded* in empirically obtained data.

Its method involves the development of theory inductively by a researcher through a process of comparative coding and analysis designed to identify conceptual categories within and between data, which are 'theoretically sampled' so that new data sources inform the emerging theory. This process continues until no further conceptual categories can be derived and theory is developed (Glaser and Strauss, 1967; Strauss and Corbin, 1998).

As a methodology, grounded theory requires a researcher to engage deeply with her data, recording reflections about them and responses to them, and using these reflections, in the form of memos, to inform analysis. It also enjoins 'great seriousness about the words and actions of the people studied' (Strauss and Corbin, 1998, p. 6). This serious attention to research participants places their lived experience, and the meaning-making they engage in about it, at the heart of the research process. In a later constructivist evolution of the methodology, interpretive elements are even more clearly fore-grounded: this emphasis has its pre-shadowing in earlier versions where there is a concern with actions and processes in analysis, but in the later version it is used to draw closer attention to the identification of assumptions, unspoken knowledge, and pivotal moments of change (Charmaz, 2014).

Since its inception, grounded theory has, to a certain extent, taken on a life of its own (Corley, 2015). In fact, ‘methodological dynamism’ has been identified as a distinctive internal feature of the method (Ralph, Birks and Chapman, 2015). It has evolved through the work of its original founders (Strauss and Corbin, 1998; Glaser, 1999) and through practical application by many others. It can be used in a way that includes an ‘interplay of methods’, including quantitative ones (Strauss and Corbin, 1998, p. 33). It has been applied within critically oriented studies (for references to a range of these studies, see Charmaz, 2014, 2017). While it is commonly cited in research across many disciplines, it has been particularly criticised where the term has become a shorthand for the application of a set of coding procedures tending to lead to complex description rather than to theory (Braun and Clarke, 2006) or where theories are thought to have been ‘forced’ rather than allowed to ‘emerge’ (Glaser, 1992).

3.5.2 Critical realism and grounded theory

The question of whether grounded theory and critical realism can be combined to create a research methodology has been explored by researchers in various disciplines, including human geography (Yeung, 1997; Hoddy, 2019), social work research (Oliver, 2011), leadership studies (Kempster and Parry, 2011), and organisation studies (Belfrage and Hauf, 2017). The analytical methods of grounded theory have been suggested as comparable with the categorising and connecting strategies deemed a necessary part of a realist approach to qualitative research (Maxwell, 2012, pp. 119–120). A validation model for a philosophically realist version of grounded theory has been proposed and tested in an empirical study (Lo, 2014). It is worth noting however that grounded theory was considered, but then rejected as a suitable method for use within a critical realist exploration of the causal explanations for particular kinds of health-seeking behaviours among HIV-infected South Africans. This was because of specific misgivings about the method's ability to fully address the 'structural constraints and discursive resources' which inform the way in which people act (Decoteau, 2016).

Applied on its own as an approach, grounded theory would have allowed me happily to address the first three elements identified above as desirable in a methodology for this study. Although it can encompass quantitative elements, it is qualitative at heart and in practice generally (Glaser, 1999). The method encourages an analysis of participants' experience as process, defining and conceptualising relationships between and within their accounts of their experiences, as well as decisions made and actions taken (Charmaz, 2014, pp. 245–6). It embraces a certain amount of reflexivity, particularly in its constructivist form. The act of memo writing, the reflectiveness this requires, and the inclusion of these reflections and researcher insights within the theory building phase, are an integral part of the grounded theory analysis process.

Additionally, the method's primary modes of analysis are inductive and abductive, rather than deductive. Induction is a type of reasoning in which concepts and theoretical patterns are extrapolated from, or 'grounded' in primary data. This contrasts with 'deduction' in which the process is reversed. Abduction is a creative process in which a researcher scrutinises inductively derived data, notes something puzzling or unexplained, considers a range of possible explanations for this finding, and then hypothesises and tests these to arrive at the one that is most plausible. Abduction has become prominent in the recent constructivist revision of grounded theory, where it is foregrounded within the processes of theoretical sampling and constant comparison (Charmaz, 2014, pp. 200–204).

Here we reach a point of differentiation between grounded theory and critical realism that forced me to think deeply about whether combining the two was appropriate.

A critical realist grounded theory would need to make a distinctive move into reasoning in a retroductive fashion (section 3.4.2). As noted above, at least one form of grounded theory already uses abduction, whose analytic movement is broadly similar, and therefore some researchers have suggested there is no reason why it should not also be able to encompass retroduction as an analytic strategy (Yeung, 1997; Oliver, 2011; Belfrage and Hauf, 2017). However, unlike abduction, retroductive analysis starts with a specific question rather than an observation of something puzzling or unexpected in the data.

Orthodox grounded theory stipulates that the researcher should ‘enter the research setting with as few predetermined ideas as possible’ so that subsequent observation and analysis is not ‘filtered through and squared with’ pre-existing hypotheses and biases (Glaser, 1978, pp. 2–3). Researchers are specifically discouraged from bringing existing theory into their grounded theory analyses. However, even in its Glaserian form, the method is not as hostile to prior knowledge as might be imagined. Glaser termed the grounded theory researcher, a systematic and ‘highly sensitized’ agent (1967, p. 251), sensitised in particular to and by the data collected in the field, but also, inevitably, by personal experience and through being ‘steeped in the literature’ of her field (1978, p. 3). The constructivist revision of the method also fully grasps the impossibility of an objectivist ‘blank slate’, recommending instead that we should take ‘the researcher’s position, privileges, perspective, and interactions into account as an inherent part of the research reality’ (Charmaz, 2014, p. 13). Finally, it has been suggested that a realist approach to grounded theory analysis might help to overcome some of the problems created by the recognition that it is neither possible, nor desirable to approach data without prior knowledge, experience or theory (Yeung, 1997).

A use of retroduction, with its distinctively iterative analytical movement and its use of prior theory, would move me methodologically further away from orthodox Glaserian grounded theory. The analytic movement required is not going to be the horizontal fracturing and then weaving ‘the fractured story together again’ coding and theoretical coding process described by Glaser (1978). Usefully, however, it does have similarities with Strauss and Corbin’s description of axial coding: ‘When analysts code axially, they look for answers to questions such as why or how come, where, when, how, and with what results, and in so doing they uncover relationships among categories’ (Strauss and Corbin, 1998, p. 127). Strauss and Corbin advocate looking for answers to such questions by closely relating structure (the things that set the stage for action and interaction) to process (the actions and interactions that occur over time and in response to certain problems and issues). Their suggested method of doing this is through the application of an organising framework (they call it a ‘paradigm’) into which a researcher can gather together and order these data relating to structure and process.

Most of those who use grounded theory, like the majority of qualitative researchers, will not wish to make causal claims, or even consider such claims an option (Lincoln and Guba, 1985; Urquhart, 2013). Strauss and Corbin put it slightly differently. After noting how causality will mean different things to those working in different disciplines or within different specialities, they write: ‘Our concern, as analysts, is not so much with causality as with conditions of various types and the way in which they crisscross to create events leading to actions/interactions. When people act, we want to know why, how come, and to what situations, problems, or issues they are responding’ (1998, p. 133). Their interest, they tell us, is in the ‘complex interweaving’ of conditions and events that provoke human responses. Additionally they state, ‘the analyst might identify changes in the original situation (if any) as a result’ (1998, p. 132). Despite this afterthought, in essence here, I see them as describing a method whose prime purpose is to engage with the process view of causality described above and which is used in this study.

My final requirement of a methodology was that it allow space for a critical voice. The grounded theory method was developed largely before the advent of critical theory. Glaser has been particularly vehement in opposition to adaptations that appear to force data in particular ideological directions, stating categorically that ‘Grounded theorists *never* should be seen as crusaders, subversives, or underminers’ (1999). He did envisage grounded theory as having directly practical application however, noting that substantive theories bringing attention to actions and processes that might not otherwise have been visible ‘can give participants in a situation a broader guide to what they already do, and perhaps help them to be more effective in doing so’ (1967, p. 248). Nevertheless, this is somewhat removed from the concerns of social justice.

The potential for bringing together critical inquiry and grounded theory has been examined by Charmaz with the conclusion that its pragmatist roots (hinted at in the quote from Glaser above) and constructivist form make it well suited to the asking of critical questions (Charmaz, 2017). Certainly, as Wuest has suggested in relation to feminist theory (1995), Strauss and Corbin’s attention to how contextual and structural influences may impact on individual agents' actions and interactions indicates that there may be a degree of consistency between grounded theory and critical epistemologies. Charmaz describes a set of emphases that might be the subject of guiding questions for a researcher embarking on grounded theory studies with a social justice aim. These include an interest in how the actions and interactions of individuals are influenced by: resources; hierarchies; ideologies, and; policies and practices (2014, p. 327). While my study is not a critical one in the sense that Charmaz is discussing and it does not have social justice as its primary goal, it does incorporate an interest in these four elements and the way in which questions about them might stimulate critical thinking about how methodologies are chosen and used in the field of arts and dementia.

3.6 Summary

Taking a critical realist informed approach to grounded theory would give me a strong theoretical background for data collection and analysis. As a methodological package, it would enable me to describe local challenges experienced by individual evaluation stakeholders as process rather than discrete events. It provided a framework in which I could theoretically describe and relate these processes to the cultures and structures affecting arts practice, dementia and dementia care contexts, the demands of policymakers, and influential discourses such as that of evidence-based practice and cultural policy. It offered an opportunity to explore the culturally and socially constructed method assemblages and hinterlands in which those involved in evaluation find themselves working. In contrast to traditional grounded theory, the approach was tempting in that it might allow me to identify causal connections and gave me permission to apply prior theory in a retroductive fashion. The critical focus aligned with my conviction that research should seek to reveal contradictions and assumptions and to provide the motivation and tools for change. It is not an approach that has been previously widely used and therefore the thesis also provided an opportunity to evaluate its success in practice (or otherwise) in achieving the aims and objectives of the study.

Chapter 3 set out the theoretical influences and philosophical framework underpinning this study. It outlined how these have been combined methodologically through use of a critical realist approach to grounded theory. In the following chapter, I describe how the aims and objectives of the study were practically accomplished, outlining its methods and design, and addressing considerations such as credibility, ethics, and reflexivity.

4 METHODS

4.1 Introduction

In this research the proposal was to explore the methodological challenges reported and experienced by evaluators in the field, as a means of signposting solutions for future evaluation practice. This has been achieved through attention to a series of objectives:

1. To conduct a literature review that identified the challenges currently reported by evaluators and researchers in the field of arts and dementia;
2. By interpreting and synthesising review findings, to develop a conceptual understanding of the nature of these challenges, any underlying or connecting themes, methodological 'hotspots' or areas that were under-described;
3. To conduct a series of interviews with evaluation stakeholders and to analyse the results with a view to identifying factors that might be constituting barriers to effective evaluation practice;
4. To use the findings from (1), (2), and (3) in signposting solutions that might be translated into practice for evaluators, arts practitioners, and those who fund and commission research.

The objectives described above were addressed through use of a hermeneutic framework informing a narrative synthesis of the literature, and a qualitative design for data collection using semi-structured interviews. The approach to analysis and theory development was derived from and inspired by a blend of critical realism and grounded theory, as has been described above (section 3.5.2).

4.2 Data collection

Qualitative, interpretive methods are considered appropriate where the aim is to make sense of the world in terms of human experience of it (Denzin and Lincoln, 2018). My chosen method for data collection in this research was the interview. Because of the potential this offers to elicit a 'first-order understanding' of the world from research participants, it aligns with the theoretical and epistemological framework for this study. A semi-structured interview form allows the researcher to play a visible role in the process of interpreting and producing knowledge (Brinkmann, 2018). It also focuses the topic under discussion, while leaving space through which the interviewee's own perspectives and insights can be voiced.

4.2.1 Interviews

Research interviewing has been criticised because of assumptions by researchers that it gives access to authentic experience (Atkinson and Silverman, 1997). It is important to acknowledge such criticisms. Charmaz (2014) has argued that the result of an interview is always a construction, or reconstruction, of reality, which will then be the subject of interpretation by the researcher. Critical realist thinking would hold that the *realities* of the world and our experience of it are always and only accessible to us through theory. These understandings have underpinned conduct and analysis of interviews in this study.

Evaluation theorists have recognised a need to account for the involvement in and influence on evaluation of multiple interests, commonly terming individuals or groups who represent these groups as 'evaluation stakeholders' (Patton, 1997; Clarke and Dawson, 1999). For arts-based activities for people living with dementia, these include: (1) evaluators and researchers; (2) artists and facilitators of arts activities, and; (3) funders and commissioners of evaluation and of arts activity, including those responsible for managing arts programmes. These groups provided initial categories for a purposive sample in this study.

Interviews were conducted in two phases. The first was directly informed by grounded theory methods and involved 21 interviews. The second phase involved a smaller sample of five, and the aims of the interview in this phase were fundamentally realist and retroductive, in that they involved inferential checking of potential categories or theories derived from analysis in the first phase.

The interview process was challenging, a balancing act between the need to elicit a coherent analytic narrative within and between participants across the study and a desire to explore individual experience. It was hard to know at first to what degree interviewees should be directed towards certain topics – for example, those topics that had emerged from my literature review – and how much their own experience and their reconstruction of it should be allowed to be the focus of discussion. In alignment with the principles of grounded theory, I tried to enable reflection on these topics, but not to allow them to focus or restrict what was discussed. As the interviews progressed, I became more confident in enabling the experience of the individual participant to remain the centre of discussion whilst probing what, in my emerging interpretation, appeared to be theoretically central ideas.

4.2.2 Phase 1 interviews: theoretical sampling

For the first phase, guided by the grey literature review, recent evaluation reports and stakeholders associated with these were identified (section 2.3.4). These were interrogated to determine an initial purposive sample. Inclusion in this sample was defined by membership of one of the three stakeholder groups and an involvement in an evaluation of arts activities for people with dementia, in the UK between 2012-2017 for which some kind of report was available and had been publicly circulated. One aim of the interviews was to be able to explore individual evaluation cases with different interviewees who might have experienced them from different perspectives. Further to this, the grey literature also suggested additional artists and evaluators with significant experience of evaluation. This process resulted in production of a longlist of 40 potential interviewees across all groups.

Grounded theory studies approach data collection, including the recruitment of participants, using theoretical sampling. This is a process that has been described as ‘controlled’ by the emerging theory (Glaser and Strauss, 1967, p. 45). Its purpose is to obtain data that can help to explicate analytical categories or to explore negative cases (Charmaz, 2014), rather than to verify facts, represent a population, or produce generalisable results. In recruiting interviewees for this study, my aim was not to define a sample that would be representative of the arts and dementia landscape in general. Instead, the intention was to explore ideas about evaluation and methodological challenge through comparison of the context, detail, and experience of evaluation stakeholders in individual cases. I therefore selected interviewees who might aid in this process.

In line with the requirements of this approach, rather than contacting all potential participants directly, recruitment was phased. From each of the evaluation reports identified as key, I aimed to recruit two or more study participants in order to look at the experience of a particular evaluation from different stakeholder viewpoints (e.g. including an evaluator and an arts practitioner or arts manager). Initial introductions, accompanied by the Participant Information Sheet, were made by email and followed up with a phone conversation or further email contact, according to participant preference. These preliminary contacts helped the participant to establish in principle whether or not they wished to take part, and me to establish whether their participation might be useful for the project. Not all responded, some declined to take part because of pressures of time, and several were found not to be appropriate because their experience was limited or had taken place prior to 2012.

Having conducted interviews with 13 stakeholders involved in five example evaluation projects, other individuals were sought, who might introduce a further variety of experience and provide more detailed exploration of particular areas of theoretical interest. I conducted 21 individual interviews in this phase in total. The majority of these (n=18) were female and all were drawn from within the UK. Their descriptions can be found in Table 7 (section 4.2.5 below).

It proved particularly hard to recruit funders and commissioners of arts activity. For example, it had been hoped that the study would include a representative of one of the national arts funding bodies, but despite lengthy correspondence, this proved not to be possible.

4.2.2.1 Including people directly affected by dementia

People directly affected by dementia are also stakeholders in evaluation. They may be activity participants, data ‘subjects’, beneficiaries of evaluation knowledge, and in a small, but growing number of cases, co-producers of that knowledge. People with dementia are increasingly included in research, and it is well understood that they are capable of expressing views, needs, and concerns. However, intensive interviewing involving this group requires considerable thought and preparation (Cridland *et al.*, 2016). Because of the theoretical approach chosen for this study, it was difficult at the outset to be certain of the questions that might be asked of study participants. Providing an interview topic guide for ethical review purposes can be problematic where the intention is to involve people whose cognitive capacities may be affected. Some areas of discussion in the interview were likely to be abstract and they would relate to experience of past events and there is evidence that some people with dementia participating in research might find such discussion uncomfortable. It was possible that an interview of this kind might be burdensome for people with dementia where their cognitive capacities were significantly affected by the condition. If they were to be included, it was therefore likely that recruitment might focus on only those people for whom such issues would not present difficulties, leaving the experience of a whole range of others unexplored. Finally, understanding the making of methodological choices was at the heart of this thesis; although people affected by dementia tend to experience the effects of methodological decisions, they are – as has been identified in the literature review – rarely involved in making them.

As a result of these considerations, and in discussion with the supervisory team for this research, it was decided that I would not interview people affected by dementia in this study. Instead an approach was chosen that emphasised the use of secondary data in understanding the nature of their involvement in evaluation and research. This had the benefit of allowing the potential consideration of a range and depth of experience that would not have been available from discussion with a small number of individual interviewees. Throughout the course of this research, I have consistently and carefully reflected on the ways in which the condition of dementia itself might have an effect on how evaluation methods are employed. This has included attention to the presence (or absence) in evaluation reporting of people living with dementia along with active reference to publications created with and by those directly affected by dementia to help researchers and evaluators, such as the guidance developed by members of the Dementia Engagement and Empowerment Project and research involving this and similar groups (e.g. The Scottish Dementia Working Group Research Sub-group, 2013; Brooks, Savitch and Gridley, 2016).

4.2.3 Phase 1 interviews: saturation

There is no agreed method for establishing when enough qualitative interviews have been conducted in a study, or for identifying an appropriate sample size. This can be a problem for qualitative researchers required to demonstrate that a sample is adequate to establish the validity of a study's findings, that it does not reflect isolated or idiosyncratic findings, or that it has not wasted research funds or participants' time as a result of using too large or too small a sample (Francis *et al.*, 2010). It has been argued that it is not appropriate to adopt saturation as a generic marker of study quality for qualitative research (O'Reilly and Parker, 2013). For thematic analysis of qualitative data as it is practiced in research more generally, it has been suggested that the term 'saturation' may sometimes be a '*rhetorical* device, rather than a considered *methodological* practice' (Braun and Clarke, 2019, p. 4, emphasis in the original).

The idea of 'theoretical saturation' is, however, central to grounded theory. Strauss and Corbin define it pragmatically thus: 'where collecting additional data seems counterproductive; the new that is discovered does not add much more to the explanation at this time. Or, as is sometimes the case, the researcher runs out of time or money, or both.' (1998, p. 136). Rather than emphasising the act of reaching a point where themes or patterns are repeating, Glaser refers to it in the sense of ‘conceptual density’:

> Saturation is not seeing the same pattern over and over again. It is the conceptualization of comparisons of these incidents which yield different properties of the pattern, until no new properties of the pattern emerge.
>
> (Glaser, 2001, p. 191)

This is quite different to the idea that data should be collected until no new themes emerge. A useful distinction between ‘hearing it all’ and ‘understanding it all’ has been identified for qualitative research more generally (Hennink, Kaiser and Marconi, 2016) and could also apply to grounded theory. Charmaz also has concerns about the use of the term because of its potential for ‘foreclosing possibilities’ and ‘constructing superficial analyses’. Urging circumspection in regard to claims that might be made when using it (2014, pp. 213–216), she suggests instead that researchers stay close to and remain willing to grapple with their data, instead of continuing to search for further patterns in new data points. She recommends this particularly at moments when analysis gets difficult: ‘When you get stuck, go back and recode earlier data and see if you define new leads’ (p.216).

While I wrote reflective notes after each interview and identified initial potential codes in the process, time pressures meant it was not always possible to fully transcribe and analyse each individual interview before conducting the next. In practice, eight interviews were conducted, transcribed, and then analysed, followed by a further five. Eight further interviews were then conducted and each one analysed before I was sufficiently confident to end the first phase of data collection. At this point I felt that I had identified a series of rich theoretical categories, which appeared sufficient without further elaboration, but which remained to be tested and refined retroductively in the following interview stage.

4.2.4 Phase 2 interviews: category refinement

After the first round of interviews, I conducted five second phase interviews with experienced evaluators and researchers. The primary aim of these was to test and refine the analytical categories derived. A further important intention was to explore connections, comparisons and differences between evaluation in the field of arts and dementia with that of dementia research and evaluation, and arts and health evaluation more generally.

Two of these interviews were with academic researchers with wide experience of dementia research and evaluation in a general sense and expert knowledge of dementia, but little or no experience of evaluating arts activities. Two were with evaluators with experience of arts and health evaluation and research, but not dementia in particular. One further interview was with an experienced facilitator and independent evaluator of arts and dementia activity. For descriptions of these, see section 4.2.5 below.

In practice, the process of analysis and writing for this thesis has been iterative and intertwined. Categories and codes were there at the end of the interview phase, but the process of putting them down on paper and discussing them in the light of other literature shaped them further. Influenced by Charmaz' injunction above, as I wrote and thought and reflected and wrote again, I constantly returned to the coding categories, reviewing them and checking them again in the context of the interview transcripts individually, and as a whole. In doing this, my process also aligned with descriptions of the need for a 'reflexive turn' in sociological research through which attention is drawn to the researcher's role in shaping the research process and product, and the requirement that this be balanced with a responsibility to represent participants' voices ethically and accurately (Mauthner and Doucet, 2003).

4.2.5 Participant descriptions

As discussed, study participants were evaluation stakeholders. They included evaluators and researchers of arts-based activities for people with dementia (n=8), artists (n=5), and managers or commissioners of evaluation and arts activity (n=8).

To preserve anonymity but to recognise humanity, all interviewees are referred to in this research with pseudonymised first names. Generic role types have also been attached to them, but these role types should not be taken as placing boundaries around their experience. Evaluators were also sometimes artists. An arts manager might have had experience of being an evaluator, or an artist. An artist might have been evaluating his own work constantly.

To refer to the evaluation projects which link some of these participants, I have also created generic project name pseudonyms. Tables 5 and 6 below record the relevant information about each participant.

Table 5: Descriptions of phase 1 interview participants

Participant	Type	Project	Description
Tanya	Evaluator	Dementia Arts Project	Academic researcher
Francis	Artist	Dementia Arts Project	Experienced arts practitioner
Camilla	Manager / evaluation commissioner	Care Organisation	Manager of arts activities for Care Organisation
Coral	Artist	Care Organisation / Rem-Arts	Experienced arts practitioner
Harriet	Evaluator	Rem-Arts	Academic researcher with a qualitative arts background
Barbara	Manager and evaluation commissioner	Rem-Arts	Manager employed by third sector organisation delivering Rem-Arts
Eleanor	Evaluator	Dementia Arts in Care	Academic researcher
Karin	Manager / evaluation commissioner	Dementia Arts in Care	Representative of arts organisation
Charlotte	Evaluator	Dementia Artists	Academic researcher
Deborah	Artist	Dementia Artists	Experienced arts practitioner
Philippa	Manager and evaluation commissioner	Dementia Artists	Manager of arts organisation
Kerry	Evaluator	Arts in Care Homes	Academic researcher
Daniel	Funder / commissioner	Arts in Care Homes	Representative of charitable organisation
Kristina	Evaluator	Various	Independent evaluator
Sam	Manager /evaluation commissioner	Various	Representative of arts organisation
Sara	Artist	Various	Experienced arts practitioner
Justin	Artist	Various	Experienced arts practitioner
Jess	Evaluator	Various	Independent evaluator and manager of programmes of arts activities
Bonnie	Evaluator	Various	Academic researcher

Participant	Type	Project	Description
Rachel	Manager / evaluation commissioner	Various	Manager of arts programme in hospital setting
Stella	Funder and commissioner	Various	Primary care health professional with dementia expertise and experience of strategic commissioning.

Table 6: Descriptions of phase 2 interview participants

Participant	Type	Description
Julia	Evaluator (arts and health)	Independent evaluator
Naomi	Researcher (arts and health)	Academic researcher (arts and health)
Martha	Evaluator (arts and dementia)	Independent evaluator
Mary	Researcher (dementia)	Academic researcher (dementia)
Rebecca	Researcher (dementia	Academic researcher (dementia)

4.2.6 Interview topic guides

All interviews were shaped by a topic guide. The creation of these also addressed the requirements of the university Ethics Review Board, providing an indication of areas that would be discussed in the interviews. While the guide provided a structure for all interviews, in practice, the focus of attention sometimes changed during the interview, to reflect the experiences and interests of individual interviewees, and in response to the ongoing analysis.

In the first phase of interviews, the structure included an initial 'settling in' question, acclimatising both participant and researcher, and gathering relevant background information. It then proceeded with more general questions establishing how the interviewee viewed evaluation and what they might find challenging about it, finishing with an exploration of issues arising from a particular identified experience of evaluation that the interviewee had indicated they were willing to discuss in more detail, if possible, prior to the interview. All interviews ended with an invitation for the participant to discuss related issues they felt might be of relevance. The intention throughout was for dialogue to remain flexible and open-ended enough to allow individual research participants to shape discussion in ways that reflected their own experience and allowed for unanticipated themes to arise, whilst ensuring that areas of importance for the researcher were addressed (Charmaz, 2014).

The guide was devised and revised following discussion with a supervisor for this thesis. Versions of it were piloted in interviews with an artist practitioner and an evaluation researcher colleague. These interviews were not included in the research. Feedback from the pilot interviews resulted in changes to wording and structure. Further minor iteration followed, relating to the forms of questions and wording that worked best, over the course of the study.

In the second phase interviews, questions were focused on refining categories of interest identified from analysis of the first phase interviews. They were directed at establishing what the interviewee felt to be the relevance of these categories in the context of their own experience. A topic guide (see Appendix 9.9.4) was produced, based around recommendations for the conduct of realist interviews

(Manzano, 2016); this indicated a range of questions from which a few would be chosen, rather than a fixed menu. A ‘teacher-learner’ cycle is distinctive to realist interviewing, with the roles of teacher and learner being interchanged between the interviewer and the interviewee during the process (Pawson, 1996). These interviews took the format of the interviewer describing a potential theoretical category, asking for the interviewer’s responses to it, and then encouraging discussion of these responses.

The title of the study and the content of some of the questions asked in all of the interviews might have led participants to frame their experiences in terms of difficulty or challenge. It is possible that this was the case. One interviewee, for example, apologised for appearing ‘negative’ in her discussion of evaluation when in fact, she said that she had very much enjoyed her experiences of conducting it, mainly because of what she saw as the potentially beneficial impacts of her work for people affected by dementia. To counter this concern, participants were asked to describe their experience, and this description was probed more deeply where they mentioned barriers, or – indeed - solutions. If it seemed appropriate, and had not already been discussed, participants were then explicitly invited to agree or disagree with the idea that it might be challenging to evaluate in this context, and then asked for the reasons behind any opinion they voiced.

4.2.7 Telephone interviews

Of the 26 interviews in this study, 22 were conducted by telephone and four were face-to-face. Although it is frequently used for quantitative surveys, the telephone is not considered an ideal medium for qualitative interviewing. There are concerns around the quality of the social encounter it creates and the rapport that can be established during a telephone conversation, along with the loss of verbal or physical cues which can communicate additional meaning (Irvine, 2010). The telephone may add a further layer to the creation of a ‘technology of the self’, another way in which research techniques can serve to shape study participant’s responses in particular ways (Gubrium and Holstein, 2003).

There are methodological limitations and practical difficulties that will be encountered when using the telephone for intensive qualitative interviews. Some of these are technical. Phone lines drop out or are of variable quality, and phones can run out of battery. Reception quality has knock-on effects for decipherability on the audio tapes when it comes to transcription. It is worth noting that transcription agencies are likely to charge more to transcribe telephone interviews as a result of concerns about the quality of the audio-files.

On two occasions in this study, interviews had to stop mid-way and then re-commence, in order to overcome technical difficulties. This may have been disruptive of the flow of conversation. It is notable however, that for two out of the three of the interviews in which this happened, the interviewee easily picked up conversation almost exactly at the point where it had left off, without prompting from the researcher.

It can also be difficult for both interviewees and interviewer to avoid distraction on the telephone. As an example, one interview conducted was with an artist practitioner. It was a very hot day and she was sitting in a vehicle in the car park of a care home, taking a break between delivering arts activity sessions. Because of the heat, she had to keep on opening and closing the vehicle's door, which then heightened traffic noise from outside. This was a distracting situation for her. Another participant repeatedly broke off the interview to hold a conversation with another person in the room.

While face to face interviews will also be affected by distractions, this experience suggests that for some people the condition of *being on the telephone* and the opportunity it offers to speak in any location, also offers a kind of permission to be simultaneously *elsewhere* as well as in the conversation. This is a potential limitation for a method in which participants are being asked to think deeply and respond carefully. While such issues would also have affected a rushed face-to-face interview, it is possible that interviewees give lower priority to a telephone interview than to a planned face-to-face meeting, making such difficulties more likely to occur. If it was evident that these kinds of distractions might have arisen during this research, participants were offered the opportunity to talk at another time. However, none wished to do so, and despite the distractions, the judgement was made that they were still able to focus and respond with appropriate depth and reflection.

A further limitation relates to the quality of personal connection (sometimes termed the ‘social encounter’) that can be established on the telephone. It was likely that some participants would be less comfortable with the technology, and that this might be more problematic if they had not previously spoken with me to establish a rapport. To overcome this therefore, where possible I spent time in initial conversation, setting the scene, establishing the structure of a connection as part of a process of ensuring that each participant understood the aims of the research.

Interviewers can also find themselves distracted on the telephone, particularly in writing notes on points to follow up on later in the interview. Despite this, I felt that the beneficial effects of being able to do this without an interviewee noticing, outweighed the negative.

Despite the potential negatives, it has been suggested that concerns around the telephone as a medium for qualitative interviews may be over-exaggerated. There may be a ‘bias’ against using the medium (Novick, 2008), that has more to do with habit and historical practice than considered thought and experience. It is acknowledged that there are advantages to its use, including – significantly – time and cost (Sturges and Hanrahan, 2004; Holt, 2010). It may also allow participants greater privacy, especially when exploring difficult issues. In the current study, the primary advantages were practical. Participants were drawn from across the UK, and neither the resources nor the time were available for travel to interview them. Additionally, using the telephone enabled the interview of participants at times and in places that were convenient to them.

It was my sense that study participants generally felt comfortable and in control of their conversation and that the telephone was an appropriate medium to use in this study.

4.3 Ethical approval and ethical practice

Ethical approval for the study was obtained from the University of Worcester Institute of Health and Society’s Ethics Review Board. All the requirements made by the Board relating to the ethical conduct of the study and the collection and storage of personal data were complied with.

Participant information sheets and consent forms were sent by email and returned either by post or with a digital signature by email. In three cases interviewees had not returned the form before the interview and in these cases where the interview was held on the telephone, the consent form was read to them and their consent recorded orally before the interview began. A completed form was subsequently signed and returned for these participants.

Templates for these forms and documents can be found in Appendices 9.6 and 9.7, along with confirmation of ethical approval at 9.8.

4.3.1 Data processing and storage

All interviews were recorded and transcribed verbatim. Audio files were removed immediately from the recording device and stored in a secure location, in accordance with the data protection policies of the University of Worcester and the stipulated ethical requirements.

I transcribed seventeen interviews myself, and because of limited time, nine were transcribed in intelligent verbatim by a professional transcription service whose contract of employment included assurances that they would comply with all necessary confidentiality and data protection requirements for the UK. Transcripts were checked for accuracy against the recordings. They were saved in digital files on a password protected laptop and at a secure online location on the university servers. All transcripts were only accessible to me and to members of my supervisory team, although pseudonymised extracts were also made available to a fellow doctoral student with whom I had a reciprocal arrangement to check and discuss analysis processes.

All transcripts were formatted and entered into NVivo 10, the qualitative data analysis software, for coding.

4.4 Data analysis

I applied a critical realist version of grounded theory in analysing the data. There is limited guidance available on how to undertake such a process. Thus, to some extent, the approach to doing so was experimental. The constructivist grounded theory approach of Charmaz (2014) resonated best with me and I chose to support the integration of this with critical realism using the example of Strauss and Corbin's *conditional/consequential matrix* when considering how to create a framework for coding. This matrix is described as an 'analytic device to stimulate analysts' thinking about the relationships between macro and micro conditions/conditions both to each other and to process' (1998, p. 181). The form of this is discussed below (section 4.4.5).

While it is often suggested that grounded theory uses an inductive or theory led approach to analysis, as Charmaz has noted, in fact theoretical sampling implies 'abductive' (inferential) reasoning, through which imaginative leaps are made and subsequently tested. The retroductive reasoning (section 3.4.2 and also see 3.5.2) of the critical realist is not too far from this position. In this study abductive techniques informed analysis of the first phase of interviews, and retroductive techniques were applied during the second phase and when developing theoretical description.

As is required by grounded theory, I attempted, as far as possible, to make the analysis process iterative, cyclical, and constant. It therefore involved the following analytic elements, although not necessarily step-wise, in linear sequence.

- Sensitising concepts (derived from the literature review)
- Interview
- Memo-writing
- Transcription, checking and re-reading
- Initial coding
- Focused coding
- 'Connecting narratives'
- Theorising and theory development

I describe below the role played by these analytic elements in this study. Additional interpretation and analysis also happened at the final writing stage, at which point the study’s findings were further situated in relation to external theory and ideas.

4.4.1 ‘Sensitising’ elements

A number of sensitising elements became points of departure during the analysis process. These included the provisional sensitising concepts derived through the categories identified in the literature review and forming what I have termed the ‘fabric of methodological challenge’. These were accompanied by the theoretical influences introduced earlier (section 3.3). None of these concepts structured the first phase interviews, although the theoretical influences were foundational in determining the approach and methods used. It was important that interview participants were allowed to describe their experience in a way that was, as far as possible, not pre-framed.

As was previously mentioned, Charmaz notes the usefulness of ‘sensitising concepts’, derived from disciplinary perspectives or prior knowledge in creating ‘points of departure’ for grounded theory research. Moving from such starting points to the data and back again, in a tentative, exploratory way – being ready to acknowledge ‘false starts’ – is an iterative process that, she argues, can open up enquiry, rather than shutting it down (2014, pp. 30–32). In grounded theory, it is crucial that prior knowledge and experience is brought to bear lightly on the inductively derived understandings of data. ‘Knowledge coupled with objectivity’ prepares a researcher to understand, and literature ‘can be used as an analytic tool if we are careful to think about it in theoretical terms’ (Strauss and Corbin, 1998, p. 47).

‘Sensitised’ is a term that elegantly describes the way in which, over the course of data collection, an interviewer might tune herself to certain constructions used in interviewees’ accounts of events and processes, without allowing these to elide other future possibilities.

As an example, in the first phase of interviews, I found myself developing an awareness for the different ways in which participants talked about how they saw the value of their role (e.g. as an arts practitioner, an evaluator, a program manager or a commissioner) in an evaluation process. This sometimes went hand in hand with emotional responses, including tensions and expressions of confusion about what the process of evaluation appeared to demand of them, as well as demonstrations of a sense of purpose and, sometimes, a need to seek validation. One of my original sensitising concepts, derived from the literature review, was ‘value’ - or the axiological dimension of methodology. Therefore, this group of responses was of particular interest, but not in a straightforward way. It alerted me to the way in which values-based ideas about knowledge might be linked to emotional responses in individuals, how they might be expressed through anxieties and tensions, and also through active practices indicating links to other kinds of values and relating to external structures and influences.

4.4.2 Memo-writing

Memo-writing has been described as a ‘pivotal’ step between data collection and writing for the grounded theorist (Charmaz, 2014). It is a fundamental part of analysis, since, for grounded theory, analysis is an ongoing process throughout the data collection phases of a study. For my research, reflective memos were written after each interview. They were also created to support the description of focused codes, as I identified them.

Writing reflective memos impacted on the process of interviewing and analysis. As interviewing proceeded, I became particularly alert to certain forms of words and repeated phrases used by the participants, suggesting that there were certain portmanteau terms whose meanings may have been taken for granted. For example, *in the moment*, used to describe the perception that arts activities have benefits that may be limited to the present time, *intangible* to describe the effects of arts activity that many people thought it impossible to assess, or *magic moments* to describe moments of extraordinary positive impact from interaction with art or an artist. It became clear that such terms might need unpacking.

I was also attentive to whether there were absences or areas where discussion was difficult. For example, I found it interesting that when asked to talk about an evaluation in which they had been involved, some participants felt the need to reach for a physical copy of the evaluation report that they had ready beside them, sometimes even reading out its aims and objectives aloud, apparently unable or unwilling to rely on their own descriptive powers or memory to do this. This might have been a facet of the medium used for interviews; most were conducted on the telephone, meaning that, serendipitously, participants were often seated at a desk and were able to consult reports while talking, perhaps as a prompt for memory. However, it also raised questions as to whether participants might be finding the language and ideas of formal evaluation difficult to incorporate into spoken discourse, or to take in in ways they found personally memorable. Issues such as these were not addressed explicitly in the interview topic guides or during the interview, arising instead through reflection in memos written immediately afterwards, thus making these thoughts available for analysis.

4.4.3 Transcription, checking and re-reading

Listening back to an interview, transcribing it, checking transcriptions and re-reading all contribute to the analysis process. As noted above, some of my interviews were professionally transcribed, but most I transcribed myself. These activities served to immerse me in the interviews, allowing me to familiarise and re-familiarise myself with their content. Again, when the transcripts had been stored in NVivo and coded, I often found it instructive to move from a list of coded extracts to reading them again in context within each interview.

4.4.4 Initial and focused coding

In establishing the methods used for coding and analysis, elements of the grounded theory approach were integral. These included my application of two primary coding stages, in which first initial and then focused codes were identified. Coding in this study involved the categorisation of segments of data in a way that summarised and accounted for their content, and also selected and sorted the data. Codes provided what Charmaz terms ‘analytic handles’ supporting the development of abstract ideas and interpretation (Charmaz, 2014, pp. 113–116). Since the eventual aim of coding in grounded theory is to develop theory, analysis doesn’t just mean coding.

Initially, coding for this study was detailed but provisional. It focused on small segments of the data and was particularly attentive to actions referred to in it. Where possible, and throughout, I coded for gerunds, rather than general topics, in order to support an insider rather than an outsider perspective on the data. Coding in this way also kept my focus on processes, thereby supporting analysis that was able to explore the kind of connections that might be helpful to have in mind when developing theory. It also meant that I was more aware of interactions between structures and the actions and responses of individuals.

A second focused coding stage was more selective and enabled the sorting, selection, and synthesis of ideas. In grounded theory, interviews are not all coded together in one way. Focused coding in later interviews builds on initial coding done in early interviews. Sometimes, this meant returning to earlier interviews to check and reformulate, or it meant using the more detailed level of coding on a later interview which felt somehow puzzling or didn't fit the schema being developed.

All my coding was conducted using NVivo10. I found the software extremely useful for linking extracts to particular codes and for establishing a traceable path. I also wrote memos in the software and stored these for analysis alongside the transcribed texts. However, as specialists in using data analysis software recognise, such tools are no substitute for the mental hard work of actually doing the analysis (Woolf and Silver, 2018). For me, this entailed a combination of the software with writing, reading, and then also more material and visual means of representing the codes and categories I identified – writing post-it notes and arranging and re-arranging them on a wall, for example.

4.4.5 Connecting narratives and retroduction

For this study, the final stage of analysis was to move from focused coding to development of theoretical categories, and finally a theoretical description that might answer the research question. It was at this stage that I applied methods more directly inspired by critical realism. In particular, I developed what I have called 'connecting narratives'. It has been suggested that a critical realist will have a keen interest in what connects an abstracted concept to the originally observed phenomenon that gave rise to it *within the context* in which it originally occurred, and less concern for comparative relationships drawn *between* categories (Maxwell, 2012, pp. 109–117).

Inspired by Strauss and Corbin's organising frameworks, and alongside Maxwell's advice on connecting strategies, I created a particular kind of memo to support capture of the process and structure of a concept in *narrative* form and from the perspective of a particular interview or evaluation project 'case'. My aim was to keep close to the interviewee's words, to retain contextual elements, and in applying a simple narrative structure, to try to understand how the interviewee had tackled some of the methodological challenges that I felt my research was revealing. To do this I first sorted key detail into a template and then crafted this into an account that placed my interviewees' words into dialogue with my own explicatory narrative. This memo form was also useful for me in making sense of interviewees' experience within a particular case where several people had provided different perspectives on a single project. They played an important role within the study's retroductive analysis process, since they allowed me – for example – to hypothesise about why an interviewee was finding evaluation challenging, going to and from their words in doing so. They were particularly instructive in helping me think through the issue of how individual agents' motivations and ideas might have been influenced by or have influenced overarching structures and cultures.

One of these narratives is presented as an example below (section 4.4.5.1). It reflects the content of an interview with Sam, the director of an arts organisation. It tells the *story* (my interpretation, based upon her words in the interview) of her organisation's experience of evaluation, and indicates key interactions and motivations of some of those involved. It situates and uses her words in this story and clusters (Charmaz, 2014, pp. 184–186) some of the key concepts my analysis had established (for example, collaboration, reflection, quality, attitudes towards evaluation, value differences, and the uses to which evaluation is put; see Chapters 5 and 6 for findings and discussion of these).

After the coding and abstraction phases of analysis, when my attention was on seeking horizontal connections across the set of data, writing these narratives was a means to help me ensure that the codes I had were also contextualised within each individual data point.

4.4.5.1 Example of a connecting narrative

> An arts organisation specialises in bringing groups of people together in its local area. Research, evaluation and reflection form part of its working practices. Sam is its director.
>
> ‘I think it’s about knowing whether we’re making a difference and if so, is that a beneficial difference […] and if so, what is it? and if there is a difference, what is it that is making that difference happen so that we can learn from it and do again and do better.’
>
> In particular, Sam says, the organisation aims to introduce ‘feedback loops’ into its work at all levels. That means, organisationally, with collaborating organisations and the people who work in them (such as care staff), with artists and with those who take part in a project.
>
> ‘That discipline of the feedback loop and the people themselves noting progress […] People being conscious of the process and the reflection is part of what we do.’
>
> A two-year project involves artist residencies in care homes and is funded by a number of charitable foundations. The organisation is exploring the impact on residents and staff, and the issues involved in bringing high quality challenging arts practice into care homes.
>
> ‘how can you bring those particular aesthetics and those particular art-forms into a care home and work with a care home audience and not lose the essence of what the public would see if they went along to a show by that artist?’
>
> The work is evaluated internally by a contracted evaluator who observes sessions and conducts interviews with all involved. Student volunteers support, documenting sessions with film and photography. Artists keep reflective journals and encourage and gather feedback from participants during sessions, making it part of the activity:
>
> ‘…so it doesn’t feel like you’re doing evaluation, which people don’t like to do.’
>
> The questions they are interested in include: What is valuable about this kind of work? What kind of changes can be seen for residents and staff? What are the practicalities involved? What is good arts practice in this context?
>
> The residencies are all very different, involving quite different ways of engaging with participants. The care environments also vary:
>
> ‘four staff teams who behave differently, managers who are less or more supportive. All those variables that are in there.’
>
> Individual staff members may also have different ideas about what is valuable about engaging in the arts, depending on knowledge and experience.
>
> ‘So, some of the activities with the artists’ residencies that worked the best in terms of the visual arts, were more abstract […] but some people’s value of that wouldn’t be as much as if it was a clear painting of something. […] I think there’s all kinds of things in there about just valuing the joy of messing around with colours and trying them out and experimenting and seeing where it takes you which an artist will do all day and invite people to join in with, but how the population at large sees that…’
>
> The reflective process brings benefits during the project.

> ‘I think the process of doing it helped [the arts partners] while they were running the residencies in terms of that sort of feedback loop and having to check in with staff and residents all the time’.
>
> An end of project event helps to disseminate the report and what the organisation has learned.
>
> ‘I think what’s important in that is, I suppose, for me, publishing the real nitty gritty of what it takes to do this work and some of the artists’ descriptions of a particular activity or moment which really bring it to life for the person who wasn’t there.’

4.4.6 Theory development

With the narratives described above in place, alongside the initial and focused codes, it became easier to connect and link what Strauss and Corbin term *causal* and *intervening* conditions for particular concepts (1998, pp. 131–132) and to develop a theoretical explanation for my observations.

The term *theory* is one that inspires disagreement (Abend, 2008). Charmaz has described it as ‘slippery’ in the context of grounded theory (2014, p. 228). As she notes, as a result of its development, grounded theory reflects both positivist (seeking causes, looking for explanations, and emphasising generality and universality) and interpretivist (an imaginative understanding of a studied phenomenon) ideas of what constitutes a theory. In this study I define theory as a relationship between abstract ideas or concepts stated for the purpose of explanation or understanding. It is not a description of a concept but an attempt to answer questions about the world in which it plays a role: the whats, hows, and whys of it (Strauss and Corbin, 1998). But I have also taken on Charmaz’ suggestion to look on theorizing as an active *practice* (2014, p. 233), and aligned it with Sayer’s suggestion that explanation should be *critical* of the world as well as seek to represent it with greater clarity (1992, pp. 251–257).

4.4.7 Narrative graphics: Analysis blending into dissemination

Dissemination is recognised as a core research activity, usually considered as the processes involved in increasing awareness, understanding, or use of research. Funding bodies increasingly expect researchers to demonstrate the impact of their work in terms – for example – of a ‘demonstrable contribution’ to society and to the economy (UKRI, 2019). A deconstruction of the meaning of *dissemination* might challenge assumptions that research findings are a finished product rather than a partial or 'snapshot' view of the world (Barnes *et al.*, 2003).

Throughout my research for this thesis I have sought opportunities to present to a variety of audiences. I have also published one first authored paper (see Appendix 9.1 for a list and 9.2 for the published paper).

In addition to the usual peer-reviewed publications and conference presentations, more creative methods are increasingly used to make research findings visible and to support public engagement. Examples have included dramatic or poetic performance, exhibitions, cartoons, graphic novels, and – perhaps most commonly – photographic or video presentation (Kara, 2015). This goes hand in hand with an increased interest in visual research methods more generally, perhaps resulting from the growing centrality of visual images for 'symbolic and communicative activities' within contemporary culture (Rose, 2014). Graphic stories have been used in health contexts to convey powerful messages (Green and Myers, 2010). Evidence supports the suggestion that for many people the presentation of text and illustration together will support learning and may increase understanding of complex ideas (Mayer and Sims, 1994).

With these benefits in mind, and having completed the primary analysis stages of this research, I worked closely with an artist to develop novel dissemination assets to integrate graphic illustration and the words of my interviewees. This process was useful in creating images that could be used in engaging a range of audiences in the results. Quite unexpectedly however, it also introduced a secondary level of analytical process which had value in and of itself and which entailed reflexive engagement with some important ethical and theoretical questions.

4.4.7.1 A collaborative process

The process by which the graphics were produced was collaborative; the artist and I worked together, using the data collected during my research as foundation and inspiration. The artist's work more usually involves the creation of animations that tell stories supporting people to make positive health choices and engage in treatment. I chose to work with her because of the careful and sensitive way in which she had used the words of people involved in health treatments and conversations alongside visual elements and in illustration. In our project together, she introduced me to methods used to storyboard animations with clients, and we adapted these in creating a graphic narrative.

Prior to commissioning, and following a discussion of the work and its requirements, the artist had responded with a proposal for a way of working together, which we agreed. Following this, in a first conversational session, we discussed the research objectives, methods, and my findings. She visualised that conversation, using pen, paper and a pin-board. We both agreed the resulting representation to be accurate. We made a joint decision that the target audience for the graphics should be my stakeholders without access to academic publications and presentations – primarily artists, arts organisations, and the wider public.

Because of the nature of the research and its methods, I felt it was important to include and reflect the words of interviewees within the narrative. However, rather than focus on individual interviews, we used the transcripts to create three composite ‘characters’: The Artist, The Evaluator, and The Funder. These characters, as represented in the narrative, retain the differences and contradictions that exist between the accounts and experiences of individual research participants. The resulting narrative provides a reminder that experience is not universal, and that there is not necessarily one coherent story in which all evaluators think or act one way, all artists another. Each identity contains multiple perspectives – a factor that analysis methods, in seeking for coherence, may sometimes conceal.

The skeleton storyboard for the narrative was simple; it related to the aims and objectives of the study. In it, each character is briefly introduced, some of their perceptions and experience of the methodological challenges of evaluation are presented, and then extracts from interview transcripts are used in presenting a number of the solutions suggested as a result of my analysis of the interviews.

To devise the graphic narrative and following development of the skeleton storyboard, I revisited the original interview transcripts. Instead of coding them inductively, this time the storyboard elements were used as a deductive schema with set points: introducing the characters; challenges for each character type, and; potential solutions. Resonant quotes were extracted in line with this schema, sorted according to whether they were spoken by an Artist, an Evaluator, or a Funder. All were placed in tables verbatim, printed on coloured paper to indicate the character type, cut out, and then those that first the researcher and then the illustrator considered most ‘apt’ were placed in position along a storyboard line. These formed the inspiration for development of a set of illustrative roughs.

The whole process was iterative, involving both artist and researcher consulting each other and returning to the words of the interviewees frequently before a final graphic narrative was agreed.

4.4.7.2 Challenges

This was a time-consuming process for both researcher and artist, requiring sensitive and prolonged communication and collaboration to make it work. The artist was appropriately paid for the time and skills the work demanded, and this was only possible because of generous funding made available for engagement with artists as part of the grant that enabled this research. The research was new to the artist, and therefore I had to work out how best to make the work accessible for her as well as considering what elements of it could be most usefully translated for the chosen audience. There was neither the time nor the resources to address all the findings of the research. There was also no time to consider all possible audiences for the narrative, and therefore its results are limited to a certain set of perspectives. In streamlining the research findings to create a narrative, it has been simplified, with some important elements and nuances set aside.

Having made the decision that all the words in the narrative (aside from an introduction) would be those of interviewees, a series of further choices governed how those words would be presented. It was decided to identify them only by colour in the graphics, individuals anonymised as part of composite characters. Each extract was carefully scrutinised in place and certain editorial decisions were made. Wording was not changed except for the removal or addition of elisions and punctuation where this did not alter meaning in its original context. It was tempting to change words, for example where the tense was not correct, or they might have been confusing, but where this occurred, we resisted temptation or sought alternatives. This did mean that the words in the graphics were not necessarily those that either I or the artist might have chosen if we had been ‘writing the script’ ourselves. This resonates, however, with one of the goals of the dissemination project which was to highlight the perspectives of interviewees.

Another aspect of the project that demanded thought was the representation of people living with dementia. As has been previously discussed, they were not interviewed for this study, so their words were not available for use in the narrative. The literature review had shown them to be under-represented in research and evaluation about them, despite growing acknowledgement that their involvement would be both advantageous and ethical. Stigma and misunderstanding about dementia are still present, within all social and professional groups including healthcare workers, artists, and researchers.

To suggest the importance of making people with dementia active and visible as individuals who experience arts activity – rather than those to whom things are *done* – but to also reflect their continuing lack of representation and visibility, people with dementia are all individually named in the graphics, but they and these names are also perceptibly *greyed out*. Figure 2 below shows this.

4.4.7.3 Benefits

The resulting graphics have stand-alone value as crafted artefacts. They have been formatted for use in a poster and printed handouts, as a PowerPoint presentation and as individual slides within other presentations. They are available online on a personal academic website and have been disseminated through social media. It is understood that graphics can be useful in presenting research in an immediate and accessible format, and it is intended that these will continue to engage wider audiences in the results of this research and to increase its impact. The complete graphic narrative is included in the findings section of this thesis (section 5.5).

However, this was not simply a (re)presentation of research findings. Instead the work developed a kind of analytical life of its own as a result of the thoughtful methods that were used. Because final research conclusions had not been written while the narrative was being created, these methods also helped to support these conclusions (see section 7.2.5).

Figure 2: Extracted example of the graphic narrative, showing naming of people with dementia

4.5 Credibility of the research process

There is no universally accepted framework to determine the reliability of qualitative research, something that can be problematic for fields in which qualitative explorations are considered particularly useful to explicate experience (Noble and Smith, 2015). There is disagreement, even among qualitative researchers, as to what kind of considerations might be appropriate (Lincoln and Guba, 2007; Schwandt, 2007).

For conventional positivist scientific enquiry, reliability might be assessed through attention to internal validity (how truthfully the study addresses its research question), applicability (the generalisability of its results), consistency (whether its results can be replicated), and neutrality (whether it is objective or unbiased).

There is debate, however, around how appropriate it is to apply criteria derived from quantitative paradigms and approaches to qualitative research (Dixon-

Woods, Booth and Sutton, 2007). In paradigms in which it is accepted that either there is no single truth or reality (as with the philosophy that underpins constructivist grounded theory), or that if there is, it cannot be accessed except through interpretation (as in critical realism), the techniques and the idea of technical *rigour* do not easily translate. It has been argued that the concept of 'validity' is itself socially constructed (Kvale, 1995). Instead, for the assessment of qualitative studies, markers of quality may revolve around originality, appropriateness of methods and design, clarity and description of process, attention to reflexivity, and consideration of ethical issues (e.g. CASP Critical Appraisal Skills Programme, 2018).

The necessity for detailed and in-depth engagement with the phenomenon under question has also been emphasised. In their influential early work on naturalistic enquiry, Lincoln and Guba's recommendations in this area included cross-checking of data through the use of multiple sources, methods or researchers, testing and piloting of tools, hypotheses and results, and awareness of alternative explanations (1985). Later they further elaborated standards of 'authenticity', in particular the need for such inquiry to perform educative or catalytic functions (Lincoln and Guba, 2007), going on to hold the consistent position that criteria for judging 'reality' and validity should be derived from consensus around 'what is useful and what has meaning' for a particular community, as well as for any individual piece of research (Lincoln, Lynham and Guba, 2017, p. 113).

In this qualitative study efforts have been made to be transparent and thoughtful in regard to the selection and description of methods and the reporting of results. The research was designed with the intention of benefitting a range of end-users, including those in the arts and cultural industries and the health sector. Attention has been paid to how the data were collected and analysed, and to the choice of research participants, with recruitment of research participants based on a grey literature search. Each item of data has been considered carefully and respectfully, and with regard to ethical principles. Discussion before, during, and after data collection with peers and supervisors, and testing and iteration of topic guides for interviews has helped to ensure that interviews were conducted in ways that suited participants and the requirements of the study. Initial hypotheses from a first stage of interviews were reviewed with further interview participants, including experts both with and without experience of the evaluation of arts-based activities for people with dementia. Findings have been discussed with colleagues and supervisors, and presented at seminars, conferences, and events including non-academic stakeholders such as artists and health and care staff. A dissemination plan for the research has included novel visual graphics intended to engage those outside the academy in the findings, including artists and the wider public.

Further reflections on the research process, its implications for research and practice, and reflections on the methodology used, can be found in the concluding chapter of this thesis (sections 7.3, 7.4, and 7.5).

4.5.1 Reflexivity

Reflexivity ‘is the process of reflecting critically on the self as researcher’ (Lincoln, Lynham and Guba, 2017, p. 143). As discussed above, attention to this is seen as an important consideration when ensuring the quality of qualitative study. It is recognised that a researcher’s own identity, history, context and experiences will shape the way in which any study is undertaken and also influence the analysis of results. To mitigate the potential for bias as far as is possible, continued critical attention to the position of the researcher within a study is thought to be important, as is transparency about this position in reporting. In the interest of this transparency, a number of elements are worth mentioning here in relation to my background and how this may have impacted upon or influenced this study.

I have an academic background in the arts and humanities and hold a doctorate in literature. In this previous research, I looked at the way in which nineteenth-century American writers grappled with tensions resulting from the need to make money with their writing. I explored the ways in which ideas about currency and capital were reflected in the mercantile culture that printed, read and influenced their works and surrounding discourse, and how these ideas manifested themselves within a number of literary texts. As a result of this background, I come to the present study with a specific set of prior research influences and a particular sensitivity and interest in discussions of meaning-making, textual ambiguity and culture that I hope inform the work in positive ways. I am aware, however, of the need to check and balance these influences in line with the demands of the different disciplinary area in which this study is situated.

Prior to starting this research, I worked as an evaluator of arts and health projects for an arts and health consultancy. This gives me a grounding in the needs of arts organisations wishing to engage with healthcare to benefit individuals experiencing physical and mental ill-health or social inequities of many kinds. It demonstrated to me the difficulties that arts organisations face in engaging with evaluation, often because of a lack of resources, knowledge or skills. Having direct experience of evaluating arts and dementia from outside the academy gives an informative and highly practical perspective on to the present study. It also potentially opens me to bias in favour of the interests of artists and arts organisations. This, I think has been useful as a balance to the academic perspective but, again, I have needed to be mindful of it throughout this research.

Finally, I am a daughter of someone who lived, and died, with dementia. My understanding of what it feels like to care for and love an individual with dementia gives me a very human third level of experience that I have consistently brought to bear on this work, along with a determination to do it well.

4.5.1.1 The first person

In writing this thesis, I have used the first person throughout. It is conventional for research in the social sciences to be reflexive but nevertheless to strive towards objectivity in the use of language. However, nothing is ever written in a social vacuum (Mauthner and Doucet, 2003). I felt it important not to treat the process of research, analysis and reporting as decontextualized and impersonal. Writing in the first person has allowed me to acknowledge and to enable reflection on the content and process of this study. Such an approach aligns both with the subject of my research, and its theoretical influences. As sociologists and philosophers of science, Latour and Woolgar point out, it has been noted that:

> […] the "didactic dead-pan" style required of scientific reporting creates various difficulties in understanding how science is done. In particular, scientists who eschew the autobiographical form of reporting make it difficult for readers to appreciate the programme or context which provide the backdrop to reported work. […] Not only do scientists' statements create problems for historical elucidation; they also

> systematically conceal the nature of the activity which typically gives rise to their research reports. (Latour and Woolgar, 1979, p. 28)

In Chapter 4 I described and discussed the methods that have been used for data collection, analysis and dissemination in this study. I considered their strengths and their limitations in the contexts of the aims and objectives of the research. In the following chapter, I present the findings from analysis of 26 stakeholder interviews, along with the first part of a graphic narrative developed to help disseminate the work. These findings will be discussed and contextualised in the light some key concepts for the field in Chapter 6.

5 FINDINGS

5.1 Overview

This chapter presents the findings from analysis of 26 interviews with evaluation stakeholders. These included evaluators and researchers, artists, and those with responsibility for funding, commissioning and managing evaluation and/or arts-based activities. In contrast with the literature review which could only examine reported challenges, in this chapter the focus is on the experiences of those involved in evaluation. Each participant chose a particular evaluation experience or project to discuss in depth. Many were very candid about the difficulties they had faced in working in the field, as well as the joys.

These findings have been produced using an analytical process described above (section 4.4). Rather than simply describing the challenges stakeholders faced, I have sought to explore the meanings they made of these challenges as well as the way in which the work involved interaction with their surrounding cultures and structures.

Having identified initial and focused codes, I shaped these into three categories. In each there are a number of sub-themes. These are described and listed in Table 7 below.

Table 7: Findings: Category descriptions and sub-themes

Category description	Sub-themes
1. Tensions and differences expressed through **value preferences and practices**.	Ideas about knowledge Valuing the arts Valuing evaluation Advocating for Thinking about quality
2. Making **meaning** in complex and difficult situations	Fragmented communication Interpreting momentary and intangible experience Using contextual information
3. The experience of **aiming for rigour** when evaluating in arts and dementia contexts	Working with existing programmes Worrying over objectivity The burden of evaluation Recruiting participants Ethical tasks Describing and defining

5.2 Value preferences and practices

Projects discussed had been evaluated by researchers and evaluators whose methods and training reflected different disciplines and research paradigms. Evaluation stakeholders came from different sectors. They described the repercussions for evaluation of differing value preferences and practices. These were expressed through differing methods of gaining or producing knowledge, and through differing understandings around the value of participation in arts activity as well as its effects. They were sometimes linked to ideas of quality. Economic and other structural considerations affected the way in which evaluations were conducted and their results interpreted. These included effects of the need for arts organisations to advocate for their work.

5.2.1 Ideas about knowledge

In discussing specific evaluation experiences, interviewees provided examples demonstrating tensions and anxieties resulting from differences in understandings around how knowledge should be gained or how it might be used.

The term *evidence* was, for example, characterised in different ways: some interviewees talked about it as information (usually numbers) that would convince those commissioning arts activity of its beneficial effects for participants. Alternatively, they told stories using terms like *witness* and *testimony* that, they felt, conveyed truths about the experiences of those involved. While evaluation might be thought of as concerning itself with evidence, it was clear that, for artists in particular, a key aim was understanding experience.

The purpose and goals, even the general concept of *evaluation,* provoked ambivalent and often negative responses from artists and researchers from arts and humanities backgrounds. Artists described feelings of discomfort associated with evaluation activities. They criticised structured surveys and measures that had failed to capture effects in ways they felt to be meaningful or inappropriate, and they complained of an exhausting multiplicity of forms. A qualitative researcher with an arts background caricatured herself when evaluating as clothed in a 'white coat', detached from her subjects. Coldness or distance were common metaphors, with one researcher talking of feeling like 'a tool', devoid of creative purpose. Conversely, among those from more clinical or health-related backgrounds, discussion of evaluation of arts and dementia was thematically positive, with one researcher referring to it as 'absolutely food for the soul' (Bonnie, academic researcher).

Artists sometimes perceived a disparity between the methods and results of evaluation and their own experiences of engagement in, and facilitation of, arts activity itself. They might characterise evaluation as 'other' or incidental to their arts practice, or frame it as a 'necessary evil', something *done* at the end of a session (a form filled out), or hurriedly completed in the car park of a care home before driving home. Some of those who had managed evaluations within arts organisations reported trying to overcome this perception, for example by making evaluation activity a contractual requirement for artists. One evaluator reported planning meetings with artists before and during a project so that she could support them in understanding what the evaluation would require.

Friction was reported in relationships between artists and evaluators, with evaluators saying they felt they might be asking artists to do things they would resent, for which they might not receive sufficient financial recompense, and which would take time out of their busy schedules. One evaluator imagined herself the 'bogeyman'. Another (Kerry) reported: 'there have been a few times when we've got together and it's felt sort of like I was the enemy'. She commented on what she perceived as artists' feelings of resentment and anxiety at being required to make their work accountable, when it was not always clear how this should be done:

> I think with the arts, it just has that special effect that no one can quite put their finger on how and why. And so that, in itself is really difficult. How do you show that and evaluate that? And some artist practitioners feel why should they have to?
>
> (Kerry, academic researcher)

These characterisations suggest that evaluation as currently conducted may involve those from arts backgrounds in ways of working that jar, or are not always aligned, with their individual or institutional value preferences.

Such misalignments had practical effects on projects. For example, clients and evaluators sometimes had different expectations regarding the aims and objectives of the evaluation:

> […] at the beginning of the project they were so sure that what they were doing was right, as it were, and what they wanted was the proof that it was right, that they were very… they were reluctant to say *'well, you know…'*. I did have one conversation where I kind of tested this out and they kind of looked at me as if, you they said *'no, no, what we do is…'* – you know – they said, *'no, no what we do is really good'*. I said *'yes, yes I know it is, but – you know – I'm not saying it's not good, but what about if you think of it this way…'*. And they – it just wasn't going anywhere.
>
> (Harriet, academic researcher)

The researcher in this case had originally wanted to explore the mechanisms and processes behind the model used for delivering the arts activities, but reported being steered by the client towards an approach that would demonstrate their effectiveness.

The arts organisation involved in the Dementia Artists project had not worked with an academic research partner before, but did have its own evaluation practices in place:

> So, the academic research is much more, or from my experience, is very much centred around 'intervention' and successful intervention,

> whereas our work has always sort of been along this idea of, we’re about ‘interactions’, not interventions.
>
> (Philippa, arts manager)

Here the academic evaluation was again characterised as exploring the effectiveness of an intervention (this time because of funding requirements), whereas what had been valued within the organisation’s internal evaluation practice was the ability to reflect on and respond effectively to emergent situations, with practitioners. In working together, both partners had to incorporate elements of each other’s approaches, although my interviews with both the arts manager and the academic researcher involved, suggested that neither was entirely satisfied with the result.

Value differences also caused difficulties during data collection. It was noted that people value different things according to their role, discipline, or experience. A director of an arts organisation suggested that the values prevalent within the care organisations she worked with, might be limiting the capacity of those working within them to observe and reflect upon the effects of arts activity:

> […] it was really clear that the ability of the staff to think about individuals and their life and what was going on for them, and notice changes if they had a volunteer [working with residents as part of the project], and if it went well or didn’t really varies across the six [care homes], depending on what value the manager put on that kind of observation…
>
> (Sam, director of arts organisation)

Sam also noted how staff from different disciplines within a hospital (a speech therapist, an occupational therapist, nurses, and a nursing manager) might observe quite different things about a participant in an arts project, depending on their disciplinary training:

> They would all say quite subtly different things because of what they were noticing. I think that’s what people are trained to observe and value, and I think the value bit is the most important. It informs what they see, and what they tell us.
>
> (Sam, director of an arts organisation)

Ideas about knowledge expressed by participants sometimes clearly had their roots in disciplinary backgrounds: ‘I’m very tied into the positivist paradigm, I can hear that in the way I’m talking’ (Mary, academic researcher). It was understood that disciplinary training might influence whether an individual evaluator or stakeholder considered particular responses and effects to be capable of evaluation. This is exemplified in one researcher's observations about the importance of what she termed a capacity for 'deep listening', learned through the academic discipline of theatre. This capacity, she felt, could reveal a range of visual and bodily or non-verbal clues in participants taking part in arts activities, some of which might not be considered 'conventionally measurable'.

Such differences could be illuminating as well as divisive, as was demonstrated in one example in which the evaluator reported the benefits of working with clinicians on the analysis of video data from arts sessions.

5.2.2 Valuing the arts

Evaluators faced difficulties because differing values were placed upon arts participation itself. Many arts and dementia projects described were delivered in health or care settings where they were considered peripheral. Cost represents one way in which this value can be understood:

> […] an artist said to me the other day: 'I've just been talking to this Extra Care housing development and they're really really keen to have arts activity and they've got a community space and they want to programme all sorts of arts activity in there.' And I said: 'that's great, are they going to be paying for it?' 'Oh, no, no. You know, they're open to partnering on funding bids that we, the artist, might lead'. And again, that tells me that they sort of want it, but they don't value it. It's peripheral. It's secondary. It's frill.
>
> (Julia, arts and health evaluator)

Artists working in care homes pointed out that care staff and management might perceive or refer to their work as 'messing around', 'entertainment', something to help residents 'pass the time', or simply a way of making a care setting appear vibrant and busy. The way in which artists are paid, and the absence of a language through which to describe and communicate the worth of their work was felt to be contributing to a lack of visibility for arts practice:

> […] when you talk to people that aren't tuned in about this work, they think that you're just messing around and sometimes people will even say to you – '*oh, is that volunteer work?'* – you know. Well, no, I mean

> I’ve got to pay the bills […] it would be nice to have more of a vocabulary around it all I think.
>
> (Coral, arts practitioner)

Interviewees felt that the care sector lacked the resources to pay the true cost of arts activities, particularly those in which professional artists were involved. And several of those interviewed suggested that arts programmes and their effects were further de-valued when they were delivered without apparent cost by staff or volunteers, or were offered free of charge because they had been funded through external arts or third sector partners:

> I am convinced that when a care home is getting something for free, they don’t value it in the way they do if they pay, putting money towards it.
>
> (Sara, arts practitioner)

In the Dementia Artists project, in which care homes received multiple free programmes of arts activity over several years, the managing arts organisation had difficulty securing a care partner willing to commit to the programme. When it did, it received little support to encourage care staff or to involve management in the evaluation.

One experienced arts project manager described its care sector partners as ‘business-driven’, recommending that arts and evaluation partners developed a way of working that acknowledged the care sector’s underlying agendas:

> I think what drives the care sector – they’re very keen on promoting what they’re doing and they’re very keen on promoting the results. So, like the films that we’ve done and, you know, the outcomes – they’ll often post that on their websites. […] So, it’s from their business model, if they’re seen to be doing things that are really proactive with their residents, their CQC ratings will go up and, you know… so, their agenda is slightly different.
>
> (Karin, arts project manager)

Another arts manager suggested that conversations about needs at the start of projects should inform collaboration between the sectors and would support greater alignment of project and evaluation aims. She noted, however, that there were few existing models for such collaboration:

> […] realistically, because funding streams haven't been particularly supporting that way of working in the past, I think on a national scale, I think what happened [in the project being discussed] was that arts organisations – you know – got together a group of partners and people and that's, you know, what we're good at. But we didn't necessarily… and we weren't working together in that way… and we didn't necessarily understand the needs of each other at that point.
>
> (Karin, arts project manager)

5.2.3 Valuing evaluation

Evaluators and artists require resources to do their work and, as was suggested above, cost presents a practical challenge for evaluation as it does for arts and dementia project delivery. There was reportedly very little money available for evaluation. Small projects often had no budget at all, so time and money were carved out of project management allocations to enable internal evaluation. Creative methods had been found to enable academic partnerships – with money, for example, scraped back through unused parts of project budgets. Academic researchers also commonly described working more hours than they had been contracted to provide. Evaluations, including those that were larger scale and more significantly funded, might be tied to time-limited project funding, meaning that they had to fit alongside activity delivery. On many of the projects described, this had left limited opportunity for planning and consultation that could have informed development of both the arts activity and its evaluation, and there was often little time available for ethical review processes at the start.

Because budgets were constrained, some evaluators said they employed a light touch, limiting the amount of contact they had with actual project delivery, and deputising data collection and ethics processes to artists or project managers. It was noted that it was not always easy to persuade artists to attend to paperwork or other evaluation activity, and that their time was already over-stretched. This was particularly problematic since most artists were self-employed and some might not be paid for additional time spent on evaluation. All of this contributed to a perception of evaluation as a burden or chore.

Arts organisations and artists sometimes also perceived funders' requirements for evaluation as excessive, given the budget for activity delivery. In a pilot project aimed at tackling social isolation and loneliness in an urban area, Sam, the director of an arts organisation described requests from the local authority funder for elaborate monitoring and outcome measurement, including pre- and post- activity assessments at each session. She felt that these activities conflicted with her organisation's primary mission to deliver arts practice, which the funding had been awarded to enable: 'it felt like we almost needed twice the money so we could spend that all on the measuring'. She was also concerned that excessive questioning before or after sessions would have negative effects on participants.

And, finally, results of evaluations themselves were not always highly valued, aside from what they enabled the organisations or individuals involved to say about their work:

> I've had… you know, organisations and… you know, that we all know whose evaluations have been hot air or that you know have had three people and they've done a massive evaluation on it. It's in the same way that, you know artists will do one five-week project and, you, know, three years later they're still talking at events about it, and quite high profiled.
>
> (Sara, arts practitioner)

5.2.4 Advocating for

Arts managers described arts and dementia projects dependent upon access to fixed term competitive grant funding, or sustained through a patchwork of charitable, local, and national funding provision. This was seen as characteristic of the arts and health field as a whole. One researcher (Naomi) referred to the resulting 'entrepreneurial' character of some arts organisations she had worked with. She commented on their sensitivity to financial risk, desire to capitalise on investment of skills and resources, and ambitions to maximise perceptions of their impact.

Health and care partners, including residential or community care, and acute and primary care services involved in commissioning activities, were also understood to be operating under financial restrictions. A primary care dementia specialist (Stella) sitting on a regional Clinical Commissioning Group (CCG) noted that while her fellow CCG members might recognise that the arts had potential for enhancing quality of life and care, this would not necessarily translate into investment: ‘unfortunately I do have to sell it from a financial point of view because that’s how CCGs work, it’s all about finances really’.

Interviewees described arts organisations engaged in projects with multiple stated aims. While the aims of an organisation usually remained stable, the outcomes they sought to demonstrate for a piece of work through its evaluation, or the way in which these were measured, might depend on the agendas of funders:

> Well, we try to have the idea and then match the funders to the idea rather than the other way round, but I think sometimes they do influence… or the source of income does influence, not necessarily what you do but how you measure it and which bits are more valued than other bits.
>
> (Sam, director of arts organisation)

This could be problematic for an organisation since, while its core values might not change from year to year, its sources of funding could, necessitating alignment with the strategic objectives of those funding sources. One researcher commented on the ‘destabilising’ effect this could have on organisations over time, leaving them without a coherent approach to the evaluation of their work, and potentially also to its delivery.

While interviewees described a desire to contribute knowledge to the wider field through evaluation and its accompanying monitoring and audit practices, many acknowledged that for its stakeholders, evaluation’s primary purpose was often to ease access to future funding:

> You know what they wanted was something they could take, understandably, and I’m not blaming them, to commissioners that would enable them to get more funding.
>
> (Harriet, academic researcher)

Artists and arts organisation representatives reported conflicting feelings about this:

> I’ve even heard people who’ve written evaluations even talking about *‘oh, you just get it to say what you want’* – all that sort of thing. So, I’ve sort of always had this sort of… but at the same time it’s like – how do we do it? I don’t know, and I’ve always been sort of scared of blowing

> my trumpet and – which is ridiculous, because the impact that the work has on people is clear!
>
> (Sara, arts practitioner)

The pressures were thought to have led to the conduct of particular kinds of evaluation and analysis. One academic researcher reported an accompanying tendency to report 'semantic' or surface value rather than to explore alternative, negative experiences, or questions of process. An arts manager with experience of conducting her own evaluation activity, and who provided a bridge between the client organisation and the academic partner in the Rem-Arts project, commented:

> I'm to blame as well in the early days, just counting people in and *'how do you feel about this project? Was it good?'* But without any theoretical base to it. I think that's what's lacking in many organisations.
>
> (Barbara, arts manager)

5.2.5 Thinking about quality

As well as demonstrating instrumental outcomes resulting from their work, for some stakeholders the concept of evaluation was associated with understanding or demonstrating ideas of 'quality' and what these might mean for arts and dementia practice. An independent evaluator noted a lack of attention to negative experience:

> Most evaluations, perhaps so-called evaluations, much of the research that is carried out by arts organisations is in response to a grant, and it is a condition of the grant that such evaluations or research should be carried out and it's immensely difficult in those circumstances to, for an organisation to return a piece of research that says *'you know what, actually it didn't work at all'*. And for those reasons I think there is an enormous amount of, not only publication bias in this type of work, but pre-publication bias as well.
>
> (Kristina, independent evaluator)

She also linked discussions of quality to evaluation's failure to demonstrate the conditions that might lead to an outcome:

> […] we think community music is good, but we don't know how we do it, and by that I mean we're not… we don't know what are the

> conditions for creating the best, most appropriate change in people through music-making.
>
> (Kristina, independent evaluator)

Others related it to ethical considerations:

> I'm aware of bad practice we see that happens at, you know, artists are – with the respect of, with respect to them as artists, you know, get money and going to work with people who are vulnerable in ways without having any training, and whatever.
>
> (Sara, arts practitioner)

One interviewee suggested a need to have a clear separation between evaluation of impact and 'quality of practice', in order that the association between the two could be explored:

> I think that one of the problems for evaluating a sector where there's such a diverse range of approaches and art-forms and sophistications of practice, is that if the evaluation each time tends to focus on impacts then you're missing out the correlation between quality of practice and impact.
>
> (Julia, arts and health evaluator)

Aesthetic quality, (ideas of what is 'good' or 'excellent' about artistic production), was thought by some to be of particular significance in the context of dementia. This was because of a concern to avoid assumptions that people with dementia might constitute an audience that might be less discerning, or less able to make judgements of quality themselves. As one researcher (Harriet) put it, it should not be acceptable to deliver 'any old thing that happened to keep them [people with dementia] entertained'.

5.3 Making meaning

This category describes the way in which stakeholders struggled to interpret what was meaningful about the engagement of people with dementia in arts activity through its evaluation. In particular, evaluators expressed concerns over the use of existing measures and tools. These concerns were, variously, related to: difficulties in communication and cognition experienced by people living with dementia; the practical and theoretical applicability of these measures in the context of the settings in which arts and dementia activities took place, and; their suitability for capturing effects related to the ‘in the moment’ and subjective nature of the arts experience. Many of those involved in evaluation, and particularly artists, also described the effects of engagement in creative activity as ‘intangible’.

5.3.1 Fragmented communication

Involving people living with dementia in evaluation was commonly reported as challenging. Symptoms of the condition meant that evaluators had to think carefully about the evaluation tools they used and when they used them:

> […] one of the things that you can’t do is the kind of classic evaluation strategy of doing a baseline assessment by asking people a set of questions and then reflecting on them afterwards at a later stage because quite a lot of them, often the people we are working with are post-verbal or their communication is fragmented…
>
> (Harriet, academic researcher)

Problems were reported with many of the existing evaluation measures and tools.

5.3.1.1 Conducting surveys and interviews

Interviewees described projects in which those living with dementia had been required to respond to standard surveys and questionnaires. Participants, for example, may have been asked to have a brief conversation with an evaluator about what had gone on during an arts activity and to report on abstract topics such as the quality of their experience or their feelings of wellbeing or happiness. None of these were reported to have been easily achieved. Interviewees also mentioned that hearing, sight or mobility impairments that had complicated and compromised communication. Some noted ‘gatekeeping’ by caregivers convinced that there would be no point in an evaluator attempting to interact with an individual or a group of participants. Where dementia was advanced or the participants were unable to respond to or with language, some interviewees felt that communication for evaluation purposes was impossible.

Interviews with people living with more advanced symptoms of dementia, in particular those conducted outside, before, or after the arts activity, were sometimes thought to be ineffective.

> […] this project hasn't hugely focused on participant feedback in that way, interviews would be ineffective with people with dementia.
>
> (Karin, arts manager)

Interviews had also been experienced as problematic even if participants had shown increased communication capacities during an arts activity.

> There was one strand of sessions where the people in a care home have quite moderate dementia, but they're quite vocal in the sessions and afterwards. So I thought that they'd be able to speak about their experience. But when it actually came time to have an informal interview with them after, it was just so much more challenging.
>
> (Eleanor, academic researcher)

As Eleanor went on to note, it could be the case that: 'once the artist had packed up, it was kind of, that had closed'.

Evaluators had concerns over the use of tools not specifically designed for people living with dementia. One researcher involved in a project involving arts activities in a gallery, described a client's failed attempt to use Outcomes Star (Good, Lamont and Triangle Consulting Social Enterprise Ltd, 2018), an established proprietary tool designed specifically to measure and support enablement or positive change and adapted for different participant groups. The intention had been for care staff to support the older people participating, some of whom were living with dementia, to complete the tool. Staff decided that, although person-centred, the tool was not appropriate for people living with dementia, because they felt the concept of re-enablement following engagement with an arts activity was not an appropriate outcome to measure.

Some interviewees, although usually not academic researchers, described using measures including or adapted from the Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS). There was a perception that this or a similar measure might be required by, or would be more generally acceptable to, funders and commissioners. There was, however, some uneasiness over the use of this and similar measures. One independent evaluator felt that the idea of subjective wellbeing over time and in an abstract sense, unconnected to a particular activity, would not be well understood by respondents with dementia. Others reported practical difficulties applying the measures, for example, in settings where there was little privacy or if respondents needed support to complete:

> […] people were being asked questions, people with supported mental health problems, were being asked – and some of them with dementia – were being asked questions in front of other people, with people listening. And one lady had said to [the evaluator], ‘this feels like the doctors with everybody listening’, she said. And when they ran out of time occasionally, one of the care workers would fill in the questionnaire: ‘I’ll do it, I’ll do it on behalf of her’.
>
> (Deborah, arts practitioner)

When using dementia-specific tools, some evaluators reported that they had not been able to confirm a diagnosis of a dementia. Others raised concerns about a focus in some measures on negative aspects of the condition. Questions about decreased cognitive ability, isolation or loneliness, and – directed to caregivers – about the ‘burden’ of care, were thought to have the potential to be offensive or to stigmatise.

When the communication of evaluation participants was fragmented beyond a certain point, proxies might be called on to respond for a person with dementia. Some of those interviewed saw this as both unavoidable and useful, as long as the role was clearly reported. Eleanor, for example, noted that well-informed and engaged professional caregivers had effectively ‘given a voice’ to residents in their care. However, it was noted that there were problems when these proxies, usually professional or family carers, became the primary or only source of data:

> We found that, for lots of reasons – but you’ll know that – lots of people couldn’t do it [fill in a questionnaire] for themselves, so then we’re reliant on staff. And then it really became about how able those staff were to make judgements about the person and what their needs were and what situation they were in.
>
> (Sam, director of arts organisation)

> […] the carer generally filled out the form, which was a bit – you know – some of them I could see were actually asking the person with dementia the answers to those questions and some weren’t, so I guess it was a little bit skewed...
>
> (Jess, independent evaluator)

Further highlighting this problem in an evaluation report of the project referenced above, Jess noted extracted quotes as coming from ‘couples’ rather than individual participants.

5.3.1.2 Observation

Perhaps because of the kinds of difficulties described above, five of the eight evaluators interviewed had sought instead to capture, assess, or explore the experience and engagement of people with dementia using observational methods as a primary evaluation tool. It was suggested that observational techniques would minimise distress to participants and cause less disruption of the activity, maximising its potential benefits, whilst enabling a focus on what was happening ‘in the moment’. Observational measures were also praised because they allowed evaluators to analyse more abstract constructs in greater depth than would have been possible through a survey. Some used structured non-participant observation tools including DCM and the Greater Cincinnati Chapter Observational Wellbeing Scale. Others adapted tools from other contexts, including a framework for ethnographic observation and a tool originally designed for use in measuring the engagement and wellbeing of children.

Two evaluators had chosen to video sessions, analysing these recordings afterwards, in one case using clinically-trained staff from outside the main evaluation team to bring a valuable alternative perspective to the process. Both, however, noted that video analysis was time-consuming and that particular skills and knowledge were needed for producing, sharing and storing video data.

Most observational tools described relied on records taken at measured intervals throughout the activity, sometimes alongside observations before and after for comparison. Those who had used time sampling methods with video technology felt that although some important moments might be missed, video could minimise the potential for bias. Some tools allowed the evaluator to develop quantitative profiles, enabling analysis of individual progress within and across groups, and this was considered particularly useful when they could be combined with qualitative and contextually specific data from other sources, such as interviews or reflective journals. Examples included DCM used to capture individuals’ experience of interaction with their environment and others in it over time, and a measure that enabled analysis of the ebbs and flows of the processes of creative engagement within individual sessions.

There was little consensus on which tools had been most useful or effective. The comments of evaluators confirmed a sense that the suitability of existing observational tools was dependent on the evaluation context and its questions. Kerry suggested that while – in contrast to measures taken before and after – observation might be useful in capturing what happened during an arts activity session, it was of limited practical use on its own in a large-scale evaluation where one tool was unlikely to be able to assess all elements of potential importance and where evaluator resources were thinly spread. In another project, Harriet voiced concerns that DCM, with its focus on person-centred care, had not reflected the evaluation team's interest in the relationship between 'the person with dementia and the artist, the space, the whole environment of it' – something over which arts and dementia projects often seek to have an effect. Concerns were also raised by one evaluator about the 'rigid' nature of some tools that they felt lacked the capacity to assess unexpected effects of arts activity. Some art forms also made certain kinds of structured observation more or less practically difficult. For example, Eleanor found DCM straightforward with individuals taking part in a music activity (perhaps tapping their feet, picking up instruments, or singing along), but in a visual arts or crafts session requiring smaller, more individual, and more focused kinds of dexterity and concentration, reactions and interactions had been difficult for her, as an observer, to assess.

Finally, evaluators noted that observational methods had introduced ethical complications, including difficulties with their application in public places, or before and after an activity in busy dementia care settings where personal care and other routines could disrupt them:

> […] the care staff are coming in to remove people, the artist's packing up, it's quite chaotic. There's a lot of distractions going on […]
>
> (Eleanor, academic researcher)

5.3.1.3 Reflective practice and creative or arts-based methods

Challenges such as those described above led evaluators and artists interviewed towards alternative modes of evaluation. Most common amongst these were tools and methods relating to reflective practice and involving artist practitioners.

Artists had often been asked to keep reflective diaries as part of an evaluation. These recorded observations of what had gone on in sessions including their successes and failures. To standardise their content, evaluators sometimes provided templates with prompt questions for artists to complete:

> Each of the sessions, I ask them to do sort of diary sheets, and then we – latterly again, it's a sort of model we've developed – so we would ask the research question and then look at sort of pointers within that to draw out, so that artists would look at those particular… they're usually sort of social benefits […].
>
> (Karin, arts program manager)

However, one researcher termed such formats 'contrived', suggesting that – in her experience – they left artists feeling an expectation about how they should respond to prompts and a pressure to record new and interesting things when completing the same questions each time. The resulting reflections were recognised as partial; artists were only able to take in a limited set of interactions and may also have had little motivation to report negative effects. Evaluators described artists highlighting positive moments while skating over difficulties or concerns. In some project evaluations, diaries had been viewed as uncomplicated data to demonstrate how participants had benefitted from activity. In several larger projects it was reported that they contributed large amounts of rich and valuable qualitative data, but that time, skills, and resources for analysis of these were limited. Additionally, evaluators described challenges in understanding how best to capture, interpret, and then use such reflections. While it was reported that some artists enjoyed using these diaries, it was also noted that keeping them was time-consuming and that artists were sometimes left to maintain them in their own time, unpaid.

Different kinds of reflection processes also supported artists' evaluation of their own ways of working. Some described engaging in both formal and informal reflective dialogues with themselves and with others, and noted how this had enabled them to shape and adjust their practice:

> I'll, you know, get in my car and think: *'Did that work? Did that... what went wrong there? What went right there? What shall I do again?'* […]

> I don't just sort of like do it and then leave. I… I think about it a great deal...
>
> (Coral, arts practitioner)

Some reported observing and assessing participants' interaction and engagement during and after an activity. One practitioner described an active process of 'fine tuning' her practice through reflection, another talked of continuous 'small feedback' loops that informed action and might draw attention to important issues, including those with ethical implications. Tanya, an academic researcher with her own experience of arts practice noted:

> […] most of us have some form of small feedback, commentary if we don't write it, we certainly think to ourselves *'actually I wouldn't do that like that next time'*, or that maybe, you know, inducing tears needs to be carefully thought about and how we might manage that and is it appropriate in this group. And – so I think we do have that level of informal and continual evaluation in our practice. I suppose we should, we should have enough humility to really take that on board and try and adjust our practice through that.
>
> (Tanya, academic researcher)

Sam spoke of the 'reassurance' an artist could derive from reflection, associating the process with self-care and supervision:

> […] reassurance for [the artist] that they're doing the right thing, that it can be a difficult community to work with sometimes and you can feel like: *'oh, I don't know if they want me here. People aren't responding in the way I expected, so am I doing it right? Should I change what I'm doing?*'
>
> (Sam, arts organisation director)

Justin described periods of 'rehearsal' before, and reflection time after, the sessions he led, terming them spaces for preparation and for catharsis or recovery, respectively.

More interviewees reported an interest in creative evaluation methods than had actually used them. Two described simple visual or interactive tools used to capture mood or feedback from participants immediately before, after, or during activity sessions:

> Sometimes through the arts activities, actually finding a way of doing that that might be doing something with colours, holding something up, acting something out. *'What do you feel about this morning?'* You act

> out your best moment or something like that so it doesn't feel like you're doing evaluation, which people don't like to do.
>
> (Deborah, arts practitioner)

> […] we ask people when they come into the session, before they start, to choose a colour and match it with a word that describes their mood right there and then […].
>
> (Rachel, arts programme manager)

Only one academic researcher had made use of visual methods of these kinds, and had done this while working closely with artists and art therapists:

> […] so asking people if they could draw how they're feeling today. Squiggles, or a picture, or… so we use that as a wellbeing tool, and look at maybe the colours, working with music, with art therapists looking at colours and shapes and that may be sometimes before the session starts and sometimes after the session, immediately after the session.
>
> (Tanya, academic researcher)

While arts-based tools for data collection or elicitation were thought to be potentially more engaging for participants than a formal survey, or more appropriate to the content and less disturbing of the mood of a creative session, concerns were raised over how the resulting data could be analysed or interpreted:

> I often think that the phrase 'a picture is worth a thousand words' or whatever it is, is entirely the wrong way round. In fact most people who do that sort of thing, and demonstrate, say visually, a distance travelled or whatever, then need to use a thousand words to explain what the picture means.
>
> (Kristina, independent evaluator)

It was sometimes suggested that methods and tools of this kind would over-simplify complex issues, particularly if the data produced were analysed and reported simplistically. In addition, several researchers noted that while individualised and project or artist-specific methods might suit the practice of an individual organisation or artist, there was a risk that they created data that could not then be easily transferred or integrated within a wider evidence context.

5.3.2 Interpreting momentary and intangible experience

Artists in particular ascribed a quality of *intangibility* to the effects of arts experience. Many were suspicious of evaluators' abilities to access and make sense of this. Interviewees sometimes also described their sense of arts activities as being experienced by people with dementia 'in the moment', suggesting that these effects were not easily captured by standardised designs or pre and post measures. Contextual data were noted as providing a key to interpretation.

5.3.2.1 Intangibility

Stakeholders interviewed had worked on projects in which the focus had been on psychological, social, artistic, and also more functional outcomes for individual participants, such as the prevention of falls. These had also been the focus of most of the evaluations. There was, however, a common concern among artists that such evaluation was failing to capture and convey the quality of an individual's experience of arts engagement. The qualities of this experience were thought to be of particular value, although liable to being considered by others in the same breath as 'aesthetic things that are sometimes dismissed as a bit airy fairy' (Deborah).

Deborah talked about what she felt as the difficulty of documenting, reporting, and making sense of elevated or 'special' moments of experience within conventional evaluation frameworks:

> […] if I interview people about… interview artists about their sessions – what went well, what didn't go so well, what they… exactly the questions you're asking me… They'll come up with the kind of things, magic moments is a shorthand for it. There are some moments where something special happens, and it's really difficult to fit that into some of the more conventional ways of evaluating the work.
>
> (Deborah, arts practitioner)

Other outcomes, perceived by stakeholders, but not thought easily measurable, included changes to the 'feeling' or 'atmosphere' in a room; the nature and quality of relationships between people; an enhanced capacity for non-verbal communication; and the developing agency of an individual. Effects of this kind were also described with reference to terms such as ‘lighting up’ (of a face), ‘quickening’, ‘special’, or physical ‘animation’.

Study participants became notably less fluent when trying to communicate these effects in interviews. Perhaps as a result, artists often turned to narrative when doing so. One artist described this kind of storytelling as, in itself an evaluative act, although not an ‘academic’ one:

> […] when people ask me what I do, I end up talking in stories. So, I end up evaluating my work by telling a story, which is very long-winded and not particularly academic.
>
> (Coral, arts practitioner)

Coral also suggested a visual metaphor to capture the changes that might take place during a session:

> It’s like if you could… if the air could change as you go in and, yeah, the colour of the air is a sort of like a… like a grey blue, and by the time you leave it’s like this golden yellow, that would be great! But you can’t clearly see it.
>
> (Coral, arts practitioner)

In the words of some of those interviewed, the ‘airy fairy’ intangibles of one perspective are weighed against the ‘gold standard’ of the other, with its ability to attract funding and engage the traction that will pull in policymakers and commissioners:

> [evaluation is] very challenging in a good way, but it’s also frustrating because we are still forced to, we kind of put ourselves in this medicalised box and that is still the gold standard for who gets the funding, who gets the policy changes […]
>
> (Charlotte, academic researcher)

> […] it’s fine if you’re self-funded, you can be as mystical as you like. But if you’re relying on any kind of public funding, there can be no mystique. You have to say what you’re doing and why it’s worth funding.
>
> (Naomi, academic researcher)

Amongst arts practitioners and organisations, quantitative methods were sometimes perceived as difficult or challenging, while qualitative ones seemed more in tune with the arts and were seen as easier to use. However, as one researcher suggested, a resort to terms such as *intangibility* to describe the effects of the arts may reflect a ‘poor understanding of methodology’, and may be leading to qualitative tools being wrongly employed to ‘do the job of quantitative’:

> So [arts and dementia organisations and evaluators] try to use qualitative methods to establish project outcomes, you know, rather than looking at what processes are going on and what is mediating – you know – project impact. So, you end up with very forced questions that are pretty meaningless.
>
> (Naomi, academic researcher)

5.3.2.2 In the moment

Among those interviewed for this research, the term *in the moment* was most commonly used to indicate a view that engagement in the arts has transitory effects that, while considered valuable, are not easily captured using traditional research and evaluation methods: ‘The arts, it’s something you see in that moment, isn’t it?’ (Kerry, academic researcher).

Artists again often used stories to illustrate their belief that standard tools and pre / post measures were insufficient to capture or convey what was meaningful about arts engagement. For example, Deborah described a participant sitting, arms tucked into her armpits, looking at the floor throughout a dance session. As music played at the end while others left, ‘she just suddenly did the opening and closing dance that we’d all just been doing – on her own, like this little solo’.

It was sometimes suggested that effects felt in the moment were of particular relevance to people living with dementia. One researcher reflected on temporality in this context, suggesting that the quality of interaction it was possible to have with a person living with dementia differed from hour to hour or week to week, since ‘in the moment is not the same moment as in the moment an hour ago’:

> That’s just kind of individual, completely changeable. And with again, which – again – the many kinds of dementia there are… for many people, they’re very in the moment, which is great, makes it really exciting […].
>
> (Charlotte, academic researcher)

Most evaluations referenced by interviewees were of short programmes. In those that were longer, evaluators had generally chosen not to follow individual participants over the whole period. Some evaluators focused, perhaps as an alternative to longitudinal study of individual benefits, on improvements occurring in the environments of care settings and the effects upon care staff of the arts activities. A few felt that particular kinds of change over time, for example lasting ‘improvement’ or improved cognition, might not be achievable goals for people with dementia. In contrast, one longitudinal study had followed the progress of the same individual participants over the period and had been able to observe and capture what was described as growing confidence and an enhanced quality of creative engagement in its participants:

> So, what we were seeing then is the, the capacity for people with dementia to grow and develop, to rehabilitate, to enjoy, to be able to express themselves creatively, to be given permission to creatively express themselves, and to feel – and to be able to do that in a progressive way.
>
> (Tanya, academic researcher)

5.3.3 Using contextual information

Evaluators reported that socio-demographic information, clinical or health data, or details about the dementias with which participants were living, were rarely collected, or used if collected:

> [the participant] was one of those wonderful case studies that show very significant change over the period of time. Now, we didn’t control for medication. We didn’t control for acute health problems. Maybe the first time he came he had a UTI or, you know, he was unwell in some way and then gradually… or maybe his medication cocktail changed.
>
> (Tanya, academic researcher)

Clinical data were reportedly difficult to access. Family carers had sometimes provided information, but its accuracy or clarity could not always be depended on: ‘they’d say, *you know he started on the new blue tablets’* (Tanya).

However, stakeholders gave examples from their experience to show how contextual information of various kinds would have helped in interpreting the responses of people with significant cognitive and memory difficulties. One researcher described the embodied responses of a participant, a former skilled watchmaker, in an activity involving planting seeds:

> […] he wasn't wearing glasses, he was post-verbal, he picked up every tiny little speck of dust and seed and put it very carefully into this pot and somebody said to me afterwards, oh, he'd been a watchmaker. Now, you could absolutely see that sort of embodied memory of working with tiny things.
>
> (Harriet, academic researcher)

A lack of contextual knowledge might compromise understanding of both the qualitative experience of the individual participant or group, and interpretation of quantitative data collected during observation of an arts activity:

> I can think of one lady who was there one session. Well, she was just completely disengaged, and when I was chatting with the care staff, it was because she'd been up all night because she'd been ill.
>
> (Eleanor, academic researcher)

Gathering these kinds of contextual data is something artists described themselves doing as a routine part of their work. One arts practitioner reported her methods of assessing participants' engagement during an arts activity. These included spending time before a session talking to those who knew her individual participants well:

> […]so that they can describe to me what pleasure looks like, what Mary enjoys and how they know Mary enjoys that, so that I can actually look out for the same kind of indicators in my session.
>
> (Deborah, arts practitioner)

Deborah felt a responsibility to carry out such checks, because it would enable her to 'check out my version of meaningful'.

Attentiveness to context around an activity was also sometimes found useful in identifying unexpected outcomes that might otherwise have been overlooked. One researcher provided an example that had demonstrated to her that a focus on individual outcomes for participants might not be the only way to measure an activity's success:

> […] there was a visitor, you know, who was ordinarily just attending the gallery, who got very interested in what was happening and what was taking place. And so, she was observing what I was observing, so it, it was like a prompt. And then she started to elaborate about what her situation was and her interest and that then… she wanted to become involved and become a volunteer of the gallery. So I think there's context… And if you were just going in to measure somebody's quantitative wellbeing and, and not looking at the context of the activity itself, you would miss something like that.
>
> (Bonnie, academic researcher)

Further, the 'success' of an activity might sometimes be defined in and by the context of where it had taken place and for the individual concerned:

> So, like one of my current pieces of feedback we had was that – it was from an old lady and she said *I managed to stay awake this week* and that was… I know from the work that we do in care settings, that can be really really significant for someone and it's not that *I feel like a greater part of my community now because, you know, I've broken down these barriers* and all that sort of… for her, you know, a 93-year old lady in a care setting, staying awake was wonderful.
>
> (Philippa, arts manager)

5.4 Aiming for rigour

This section draws together findings on the theme of interviewees' experiences of delivering what many felt to be more technically rigorous evaluation, often with the aim of legitimising its results. Ideas of what is and is not rigorous are still largely defined through reference to a positivist paradigm, and to concepts such as reliability, validity, and generalisability (Kvale, 1995). Evaluators described their struggles to be methodologically rigorous as a result of the naturalistic settings in which arts and dementia activities were delivered and in relation to their struggles in working with existing programmes and in managing relationships with other stakeholders. They had concerns around working on time-limited programmes of activity and had experienced problems recruiting study participants. They also described difficulties in defining and describing arts activity, managing bias and ensuring objectivity, and with the demands and requirements of ethics processes.

5.4.1 Working with existing programmes

The activities described in these interviews were all delivered by arts organisations. As one researcher put it, they were not ‘designed with evaluation in mind’. Historically, the focus has been to support practice, it was suggested by one researcher, rather than to develop or change that practice through research:

> I think a lot of the historic work and push – and it makes sense – was to actually establish arts for health and cultural initiatives as important and meaningful. So, I think it’s been process-driven a lot by the artists and the galleries and museums and people, and there hasn’t been, there’s been a time-lag for the research fraternity to get on board.
>
> (Bonnie, academic researcher)

To a certain extent, therefore, the evaluation or research community may be playing ‘catch up’ with agendas driven by the needs of arts organisations.

Evaluation researchers interviewed described working with existing programmes, rather than setting up and then evaluating interventions. Evaluation was commonly described by interviewees, for example as ‘investigations into what happened’ or into ‘things that already exist’. The degree to which it was possible to design and control the activity to be evaluated formed part of interviewees’ definitions of evaluation, a key marker for its difference from research.

The identification of aims and outcomes for evaluation sometimes pre-dated evaluator involvement. It was sometimes informed by the parameters of a funding call or an arts organisation’s response to it rather than the recommendations of an evaluator. Pre-determined outcomes and a lack of input into project implementation left some from academic research backgrounds with a sense of unease. Charlotte commented: ‘you’re just sort of a tool in the… you’re not doing any of the creative process. You’re not engaged in even what the questions are, what it is that’s being explored’. Harriet described the work of her team on an evaluation as doing ‘the question that you’re asked to do, not the one you would have done’. There was a sense that a lack of oversight into how activities or interventions were being implemented could be leaving important areas unexplored:

> […] if you’re so focused on just looking at those outcomes, there might be other things that are missed […] I don’t know, the context around the programme for what’s worked and what hasn’t worked and why.
>
> (Eleanor, academic researcher)

There was interest in making evaluations more methodologically rigorous, but evaluators reported limited opportunities for this as a result of limited influence over the shape or delivery of programmes. They described having little input into who attended sessions or what an activity comprised. This had led to difficulties, for example, in conducting and comparing structured observations when sessions of an art programme were delivered without consistent pattern, to differently sized and composed groups of individuals, and involving a multiplicity of artforms and facilitation styles. In summary, as one researcher (Naomi) put it, ‘because it’s a very pragmatic area, and the projects aren’t designed with evaluation in mind, you end up with this kind of, quite a messy picture’.

However, there were also concerns that developing an intervention so that it could fit a particular methodological approach for evaluation might contort the arts activity out of shape, have ethical implications, or affect its perceived or actual value to stakeholders:

> I think what’s happened is that ‘psychosocial research’ – whatever that means, let’s park that for a minute – but, psychosocial research using the methods and informed by, whatever this word means but – the epistemology of ‘positivist’ drug trials… and we’ve tried to, because they seem like the big brother or the big sister, that culture, a way of finding out, has been the gold standard as we call it for commissioning and for policy influence. Then we’ve contorted our, we’ve contorted psychosocial interventions to fit into that mode…
>
> (Mary, academic researcher)

> […] you bring a team of researchers to an existing project and they start to devise an evaluation research study around it and the danger is that they could end up changing elements of that programme. So, the idea is to ultimately increase the amounts of arts that is available to people with

> dementia, but you could actually end up reducing it because it doesn't fit with the research methodologies…
>
> (Naomi, academic researcher)

> […] you don't want the evaluation to get in the way of the activity – that's the kind of thing to remember. […] you don't want that to get in the way of making it a wonderful project.
>
> (Kerry, academic researcher)

5.4.2 Worrying over objectivity

When an external evaluator was involved on a project, stakeholders had certain expectations around their skills and the resulting enhanced objectivity of study results. But the role of external evaluator was not always an easy one and relationships could be difficult to handle. Those interviewed had experienced difficulties when external evaluators were brought in late to a study, not consulted about important project design or delivery decisions, or not introduced to key personnel. While it was thought desirable to bring an evaluator into a project early, developing useful relationships between them and other stakeholders was recognised to be time and resource intensive.

Evaluators described walking a difficult line between the funding and commissioning context in which evaluation is delivered, the content and processes of arts activities themselves, and a general expectation that they were required to take up a neutral position. Many recognised that being an external evaluator did not necessarily mean being free from the influence of others:

> […] if you're coming in as a third party, it should be kind of objective, whereas sometimes I think it's a challenge not to get steered towards certain things.
>
> (Eleanor, academic researcher).

In contrast, there might be methodological positives to having an evaluator familiar with the individuals, activities, and settings involved in a project. Tanya described the 'very pure' and 'human perspective' provided by a dancer colleague around the reactions of project participants taking part in a dance activity – a perspective she characterised as not skewed by clinical knowledge, as her own might have been. The involvement of the artists facilitating activities in an evaluation was highlighted as useful, since they might have developed a relationship with participants in their session and be able to support more useful communication. However, it was recognised that trust and familiarity were a double-edged sword; several interviewees raised concerns that a relationship developed with an evaluator or arts practitioner over the course of a project might have biased respondents' responses to evaluation questions.

The presence of an external team had also been found to create barriers between evaluation participants – artist practitioners and care staff – and arts organisations. One arts manager felt her organisation had been disenfranchised by such a relationship, unable to access and act on knowledge gained over the course of the project:

> […] we're used to talking to people, you know, our artists will ring us up after sessions. We've got a really good relationship with them and I think actually having an external evaluator really changed, sort of it gave us almost a gatekeeper and we didn't ever find out what was said at a lot of those [feedback and focus group] sessions and so I wasn't able to really improve anything or develop…
>
> (Philippa, arts manager)

On occasions external evaluation partnerships had highlighted 'cultural differences' between arts and academic partners. These had led to misunderstandings around protocol and language, including what and how claims were made when presenting findings, and who presented those findings.

5.4.3 The burden of evaluation

Evaluation activity was sometimes described as actually and perceptually out of proportion to the scale of a project. Terms such as 'tedious', 'over-egged', 'heavy-handed' and 'burdensome' were used.

> It's very easy as a researcher, or as someone leading the evaluation that you get so excited about everything that you have… all the evidence that you can gather, that you forget about how much of a burden it can be on people who aren't used to dealing with things like that.
>
> (Kerry, academic researcher)

Stakeholders described situations in which artists without specific evaluation knowledge or training had been asked to complete evaluation tasks:

> […] the artists delivering sessions like this are usually self-employed, so to ask them to be doing so much out of the sessions as well, is asking a lot. And things that we take for granted, that we know about, the artist might not necessarily understand.
>
> (Kerry, academic researcher)

However, while most artists reported little doubt that their work was beneficial for participants – often testifying to this in story form – they understood the need to demonstrate the benefits to others. One artist made a distinction between demonstrating the effectiveness of 'practice' – what goes on in a session – and that of the intervention as a whole:

> […] one knows one's practice works because one sees people's smiling faces at the end of the session and improvement in wellbeing generally. But in the times that we're in, we require evidence, scientific evidence, in order to at least demonstrate, if not prove, but demonstrate that the actual intervention works […]
>
> (Bonnie, arts manager)

While the nature of the relationship and interactions between an evaluator and an artist practitioner was often shaped by the chosen evaluation approach, the personal characteristics of evaluators also played a role. An artist whose work was the subject of structured observations noted how the 'presence' and observational styles of different researchers affected her:

> I found [Researcher A's] presence quite foreboding, whereas [Researchers B & C] were a bit softer and they, they sort of joined in a little bit, just with their facial expressions and stuff like that, you know, they were, they were less of a presence in the room. But they also worked as a bit of a focus as well so… because they were so interested, it, it kind of makes other people more interested too, so that raises the concentration levels. […] But with [Researcher A] I found that quite intimidating…
>
> (Coral, arts practitioner)

Coral had little interest in the observational method used in this evaluation: 'I just thought that they were looking for, my understanding is, how many smiles… they were sort of a grading of smiles, you know…'. Feeling herself the subject of observation, her comments suggest she also felt distanced from the evaluation and its findings. Other artists talked about the frustrations of having to use 'somebody else's tools' when carrying out evaluation tasks.

It is not only artists that may be burdened during evaluation. Interviewees described care and clinical staff being called on to provide or collect data, and sometimes also to support ethics processes. Staffing and shift patterns in clinical and care settings could make it difficult for evaluators and artists to form consistent relationships with staff: ‘people are working different shifts and all the care might be short staffed or, you know, any number of things might happen’ (Justin, arts practitioner). Tanya described several frustrating days spent on a hospital ward, during which she was unable to speak to any of the nursing staff she had hoped to interview.

The work of health and care professionals was viewed as demanding, and participation in evaluation added to these demands, particularly if there was little understanding about why it was being carried out:

> […] there was forms they had to fill out every week for each person. It was quite a heavy… in terms of what you were expecting the care staff… And it didn’t really work to be honest. The care staff just didn’t have the time to do it, or they just didn’t understand it.
>
> (Jess, independent evaluator)

One artist suggested that emotional and psychological barriers might prevent care staff becoming involved in arts and dementia evaluation activity. Arts activities might serve as a powerful reminder for a care worker of individual humanity and personal value: it ‘frees them a bit more from the role of paid carer to be able to be a human being alongside other human beings’ (Justin, arts practitioner). However, asking for evaluation and reflection on their role as carers, their interactions with residents, and the effects of arts activities on the lives of those residents, risked creating a conflict with their professional personae:

> I think it’s asking quite a lot of care staff who are, you know, working really long hours in – whatever – emotionally demanding and draining – to then also be prepared to share about that as well, and not kind of keep the professional lock down on all that stuff.
>
> (Justin, arts practitioner)

Evaluators also encountered routines and systems within care and other settings, in which personal care, clinical needs, and safety sometimes took priority over other activities. Hierarchical staff structures meant care workers could be unwilling to comment on factors that they might see as outside the parameters of their roles, as paraphrased here ' "*I'm not involved in that, I don't know, ask somebody else*" ' (Eleanor, academic researcher). One interviewee (Karin) even suggested that 'there isn't a culture in the UK that care staff actually get engaged and involved with residents in that way'.

5.4.4 Recruiting participants

Arts and dementia projects often reach small numbers of participants. While some of the multi-sited projects referenced for this research, such as Rem-Arts, Dementia Arts in Care and Arts in Care Homes, involved larger numbers, they also introduced more difficulties for evaluators in terms of the varied characteristics of individuals in the participant groups, the kinds of activities being delivered, and the differing cultures and environments of the settings.

Most researchers and evaluators described small sample size as a limiting factor, regardless of methods used. While Tanya had hoped to investigate the impact of a two-year arts project on professional caregivers using validated scales measuring burden and stress, too few caregivers saw the project through from start to finish to make it useful to do more than comment on results in the final report. On the Dementia Arts in Care project it had been planned that care staff would complete the Attitude to Dementia Scale (ADS), but too few responses were recorded to make analysis worthwhile. Several academic researchers also noted that small participant numbers, in qualitative or quantitative studies alike, had made it difficult when attempting to publish findings in academic journals.

Recruitment was routinely described as challenging, both practically and ethically. Interviewees described samples determined by who attended the activity and was willing or able to participate in an evaluation, factors over which evaluators had little or no control. An evaluator conducting observations in care homes noted that 'one strand of sessions I was evaluating, it was almost a different group every week'. This resulted in missing data-points for observations and issues around gaining ethical approval for pre- and post-session observation, since the evaluator could never be certain who would be attending and who would be observed. This kind of variability was also described as compromising accurate assessment of the effect of activity on individuals over time, since the quality of their engagement with the activity might also be affected:

> I think it takes time for residents, particularly people with dementia, to kind of get used to something new. So, it might take a few sessions for them to kind of get to know the artists. And so, perhaps if you're just seeing a snapshot of one session it… they might not be fully engaging with the activity.
>
> (Eleanor, academic researcher)

Tensions were reported between a project's need to attract and to maximise benefits for people taking part in the arts activities, and the needs of an evaluator to have some control over who was participating in an evaluation. For independent evaluators, these difficulties could be particularly difficult to manage. One evaluator managed and evaluated a project delivered across several museum and community settings involving the use of objects from a museum collection and a storyteller. A call for participants resulted in a range of people attending sessions, some living independently with their dementia at an early stage, some whose condition was advanced, who were living in residential care and unable to communicate verbally, and some with learning disabilities and dementia. Such variability created a considerable challenge for everyone involved, as both arts activities and evaluation materials had to be accessible and appropriate to all.

For projects taking place in health and care settings and involving people living with dementia, good relationships with health and care workers were particularly important. The judgement of professional caregivers sometimes determined who took part in an arts activity. In care settings staff might ensure that participants came to sessions regularly and received practical support to enable them to engage, regardless of physical or cognitive ability. But staff might also act as gatekeepers for those in their care: one evaluator (Eleanor) paraphrased a response to questions about the appropriateness of an arts project for residents: 'well, ours have got dementia, what are they gonna do?'.

It was also noted that arts projects tended to attract particular kinds of participants – perhaps those with a history of arts or cultural participation – and would fail to attract those who felt art was not for them. Illness, frailty, and cognitive decline within the participant group, as well as mortality, where participants were nearing the end of life, combined to add challenging underlying conditions affecting recruitment and maintenance of a group over the course of an evaluation.

5.4.5 Ethical processes

Evaluation involving any potentially vulnerable population has the potential to cause emotional distress to both participant and evaluator. Ethical processes, formal or informal, are particularly important in supporting interaction with individuals lacking the capacity to give informed consent. In the UK, ethics review processes govern the conduct of research, including some evaluation research, since while service evaluation may not require formal ethics review, it can be necessary in certain clinical settings. Several researchers also reported it to be essential where there were ambitions to disseminate the findings of evaluation in academic journals. While researchers interviewed for this study accepted and understood that formal processes of ethical review were sometimes necessary, their application within evaluation was not always straightforward.

Specific expertise and experience were thought a requirement for successful negotiation of ethical processes. One interviewee, managing an arts programme in an acute hospital setting and responsible for its evaluation, described the formal NHS ethics process as time-consuming and daunting:

> I have attempted to apply for ethics approval before now and actually given up because the whole IRAS thing and writing 10s of thousands of words about project proposals and research proposals – it just floored me.
>
> (Rachel, arts programme manager)

Sometimes multiple layers of ethics review were required. In the case of the Dementia Arts Project evaluated by Tanya these included national NHS approval, approval through two individual participating NHS settings, and also a set of university ethics. In the end she felt it ‘was as robust and hardy as anything you can read in a good science journal’, but the process of agreement took six months. Whereas ethics review was familiar to academic evaluators, they commented that sufficient time for it might not be planned into a project by arts organisations.

Where projects were using mixed methods, ethics could be further complicated. In the Rem-Arts project, observations using DCM could not begin before ethics clearance was obtained. Since the client organisation had received time-limited funding to deliver workshops for people with dementia rather than to conduct evaluation, they felt obliged not to wait on this until the quantitative evaluation could begin. However, some arts-based research requiring different ethical approval was able to start earlier. As a result, the researcher (Harriet) felt, the two different methodological approaches chosen for the evaluation could not be fully integrated and they felt ‘always and forever out of kilter’.

Both the terminology and the formalised nature of standard ethics documentation were thought to have created barriers between an evaluator and potential study participants. The process had been particularly problematic where the evaluator was not on hand to explain the forms:

> […] because of it being such a large-scale project, I wasn’t the one sort of gathering the evidence, if you like. I was relying on others to do that for me. So, I wasn’t there to be explaining the consent forms, although – obviously we had sort of sessions with the artist practitioners and, explaining it all – it’s all still in a different language.
>
> (Kerry, academic researcher)

Kerry reported that long-winded ethics forms had been ‘putting people off taking part in the project’. And another researcher also reported the ethics process having a negative effect on carefully established relationships:

> […] there were lots of things that perhaps weren’t relevant. And it was just, I found that such a huge barrier, immediately. Like, I suppose, but I’d spent such a long time building the relationships, and then I’d presented [care staff] with this information sheet and they were just like – *I don’t understand it.*
>
> (Eleanor, academic researcher)

It is likely to be particularly challenging to negotiate ethical requirements for the use of non-standard methods. For the Dementia Arts Project in a hospital setting, sessions were filmed. The researcher (Tanya) followed strict processes and requested signed consent in situ, at key points and at each session when an observer introduced themselves, or when pointing out the cameras to participants. Consent was requested again to confirm the use of the film for evaluation purposes. In retrospect, Tanya described this as ‘distinctly over-egged’ and said that she now made sure that consent forms were simple, ideally only requiring a one-off signature. One interviewee suggested that documentation designed primarily for clinical or medical research studies would benefit from translation or adaptation in contexts and with interventions where the potential for harm was different, as well as with people affected by dementia who had particular communication needs.

A nominated expert is sometimes used to advise on participation by people assessed as lacking ‘capacity’ if no personal consultee is available. In the Dementia Arts in Care project, the evaluator (Eleanor) felt uneasy about the clinically professional ‘snapshot’ taken of each care home resident by the nominated healthcare professional consultee because their advice and assessment were given outside the specific context of participation in the arts activity. She also identified a tension between the expert assessment and what she felt was her own obligation to consider the views of all those who might be involved: ‘the manager’s view, the family – if you can even get in touch with the family – and the resident’.

General ethical concerns were raised about ways in which the design, aims, or conduct of an evaluation might impact negatively on the experience of participants. For example, Tanya described attempts to standardise a singing intervention or groups taking part in a large research study. The groups were constituted differently, including individuals with different levels of musical experience and cognitive function. Some groups grasped the material very quickly, other groups found some of it unappealing. It would have been ethically difficult, she felt, to proceed with a standardised intervention that did not fulfil the needs of participants, even if doing so might have made evaluation easier.

> […] they would have been bored if we hadn’t added some form of harmony and then we’d come across songs that the whole group would kind of sigh and say *do we have to?* And we wouldn’t inflict that upon people, so it was difficult to standardise the intervention.
>
> (Tanya, academic researcher)

There were some moments when approaching participants for consent was particularly difficult. For example, while consent from caregivers for their family member to take part in an evaluation was needed, it was difficult to justify approaching them for it when they were distressed as a result of hospital admission:

> They were really, the carers were really… it, it… there were several that I didn't even approach because it was the wrong time to, it was the wrong time to be doing it. They just needed their arms round them really, just comforting.
>
> (Tanya, academic researcher)

Finally, in projects in which the ultimate beneficiary lacked the capacity to articulate what the activity had meant to them, and particularly where the artist had been involved in evaluation, one independent evaluator (Kristina) raised the issue of power relationships. She compared the balance of power in such a case, to that between a doctor and patient, suggesting that the ethical repercussions of this may be underestimated.

5.4.6 Describing and defining

While the experience of engaging in arts activities is sometimes described as 'intangible', the activities themselves were sometimes also felt to resist definition and description. It can be problematic for data collection if an arts activity's purposes, intent, structure, and intended effects are open to ambiguity or are liable to being understood differently by those involved. This becomes even more problematic if such ambiguity is viewed as integral to the activity or even seen as a marker of quality:

> […] the kind of flexibility which I think is one of the critical sort of concepts of the whole thing, that… the possibility of reacting fast to the unexpected, which is always round the corner.
>
> (Francis, arts practitioner)

The use of multiple artforms within a single session of activity was a feature of several projects, with artists sometimes expected to be able to move between different creative modalities, or to collaborate with others in order to do so.

When the organisation behind the Rem-Arts program commissioned an evaluation, the evaluator said, 'they used to talk about *the model* and to a certain extent, they still do talk about *the model*'. The model described was an approach the client wished to translate into an activity that care staff could also be trained to deliver. What perhaps hadn't been considered, the evaluator of the project suggested, was that while artists paired and used different artforms in complementary ways on the programme, they often worked in an instinctive way. There was little opportunity within the evaluation to explore how these combinations and the processes involved in them might impact in different ways on the culture of care in the care home, or indeed on individuals within that care setting. In fact, it emerged that the care homes themselves – along with their environmental constraints – might have been responsible for determining the kind of activity they received:

> […] we went to different units and offered them a sort of menu of different art forms. And some places were better suited to visual arts, some places were suited to drama because they had more space […].
>
> (Barbara, arts manager)

The resulting differences complicated comparisons between evaluated sessions which had already been complicated because of differences between participants and participant groups. This example also raised the question of why a care home might be thought 'suited' to one kind of session or another. Although this programme was considered by its proponents to involve a 'model', its inbuilt variability made it very difficult for the evaluators to theorise around the mechanics of the practice or to explore their inter-relations.

For some researchers interested in increasing the legitimacy of their studies, transparent project description in evaluation was 'the bottom line', with its absence seen as hampering synthesis of results across the field. Naomi recognised as a contributory factor in this failure, artists' resistance to their practice being 'known' or made replicable. She also aligned this with organisations' desire to protect their competitive edge.

> So that would be the bottom line really, for everyone to just describe their project in a transparent way. Not everyone wants to do that, because the thinking behind that is to make your project replicable. Not everyone wants their projects to be replicable because it's a competitive world full of small groups thinking that they're doing something unique.
>
> (Naomi, academic researcher)

Mary, a dementia researcher, described turning down the opportunity to evaluate a dance for dementia project because she felt it lacked clear articulation: 'What is this thing that may be having an effect on something else?'

> […] because it changes all the time, it depends on the mood of the place, it depends on who comes, it depends… it doesn't seem to be one thing. But obviously it isn't hot dinners, it is something that isn't hot dinners!'
>
> (Mary, dementia researcher)

For Mary the lack of a satisfactory 'operational definition' (an account of the procedures involved that might make the activity amenable to measurement) represented a key challenge for its evaluation. However, she suggested that this difficulty with definition was not confined to the arts. She related it to research into care practice involving people living with dementia more widely, and to the nature of caring itself:

> I think it could be almost that the very best practice in care for people with dementia is more – is probably practice that is the least well prescribed or described or defined because, by definition, it's more organic, spontaneous and creative, in the moment.
>
> (Mary, dementia researcher)

5.5 The graphic narrative

The images that follow were developed in collaboration with an artist, using a process described in section 4.4.7. They were created as a means of communicating to an audience of arts practitioners and the general public the findings of this study. In the form of what we term a 'graphic narrative', they introduce the study's subject and the composite characters of The Artist, The Evaluator, and The Funder. They then depict, using the directly extracted words of those interviewed, some of the challenges these characters faced in evaluating. The themes described in the chapter above are reflected in this narrative.

Figure 3: Graphic narrative (Findings)

FUNDER

It is crucial that arts organisations do in depth research and evaluation of the work that we're doing because all of those outcomes are going to be lost.

It gives us leverage to pursue ongoing work. unfortunately I do have to sell it from a financial point of view.

ARTIST

Our aim is for a person with dementia to have an experience of themselves as creative, as connected,

it's our job to broaden the expectations about what could be achieved.

EVALUATOR

What is the intervention?
What does it comprise?
when is the change that we are looking for to be measured?

As an evaluator, you're sort of a tool basically.

The ARTIST'S problem

What I'm most interested in is the way that we can learn how people living with dementia, even advanced stage dementia, retain a sense of self and of connection to that self and to other people.

I use an awful lot of skills that aren't necessarily recognized in the big wide world for this work… measuring atmosphere, measuring mood, measuring how engaged somebody is…

…so much of it is intangible, isn't it?

The woman who had asked me to do this session, came over and said 'the rest of them really enjoyed it, but I don't really think it's worth Jack coming next time.'

Now, Jack had sat motionless during that session with a straight face, but as I collected the ribbons in...

...he looked at me with his shiny eyes saying...

Now, that man had got a huge amount out of it, but if I'd been using ribbon waggling as an indicator of engagement, he wouldn't have been brought next week.

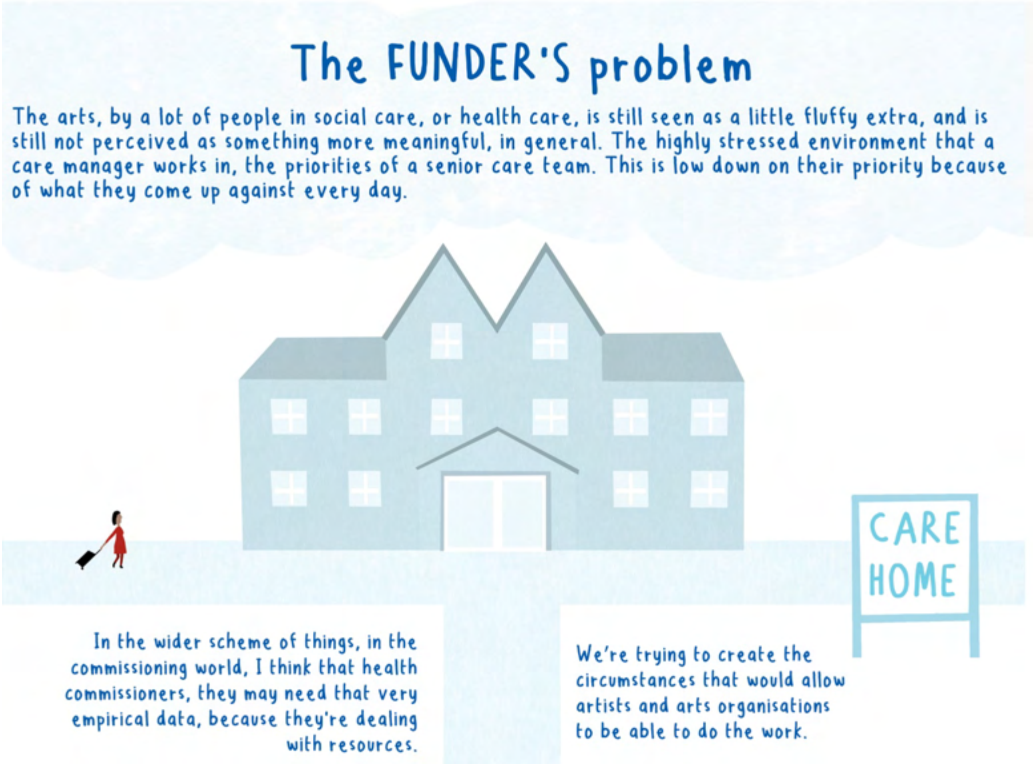
There are some moments where something special happens, and it's really difficult to fit that into some of the more conventional ways of evaluating the work.

...there's quite limited resources to do evaluations. You have to rely on the team that are there...

...a questionnaire to be delivered at the end is bonkers when we know that people have got memory problems...

...what I got was loads of data that was very hard to do anything with and I couldn't really use it.

You walk away and you think 'bloody hell, that was really hard work, but I think it...I think it was a success' but there's nobody to say 'well done'. And sometimes if you don't have that reflected back at you, you can feel that feeling of losing something in the way that people with dementia feel it...



The FUNDER'S problem

The arts, by a lot of people in social care, or health care, is still seen as a little fluffy extra, and is still not perceived as something more meaningful, in general. The highly stressed environment that a care manager works in, the priorities of a senior care team. This is low down on their priority because of what they come up against every day.

In the wider scheme of things, in the commissioning world, I think that health commissioners, they may need that very empirical data, because they're dealing with resources.

We're trying to create the circumstances that would allow artists and arts organisations to be able to do the work.

Academic research in this setting is a lot more complicated than we'd ever really expected.

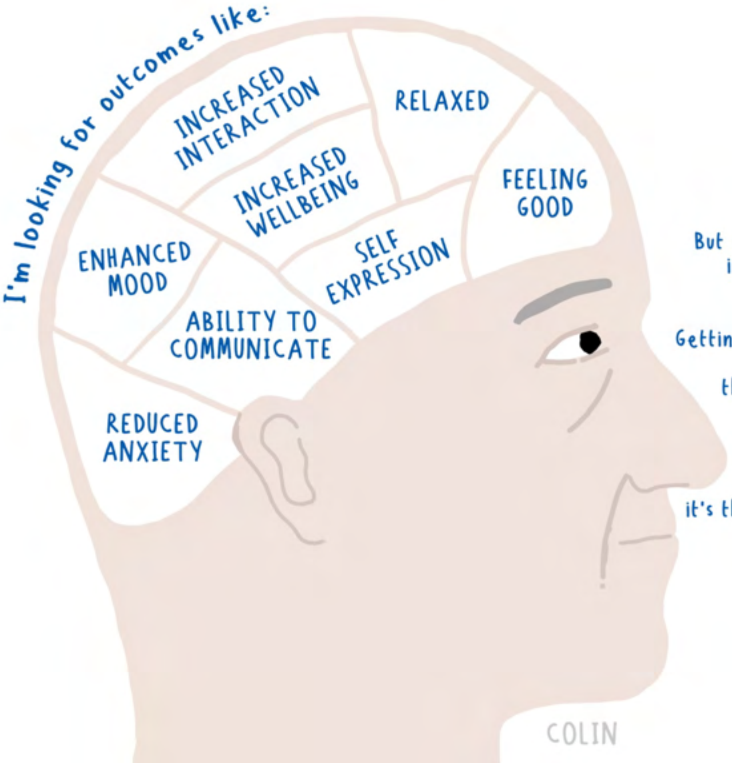
We've got our Medical Director saying:

...and I think, well, how am I going to address what he wants to hear?

The cultural divides that exist between arts and health...

...the CLINICAL is the priority.

It's just an extra level of robustness within a clinical setting, that clinical ethics approval is that much more difficult.



I'm looking for outcomes like:

But also, the other side of it
is to integrate it into the
practice of the home.

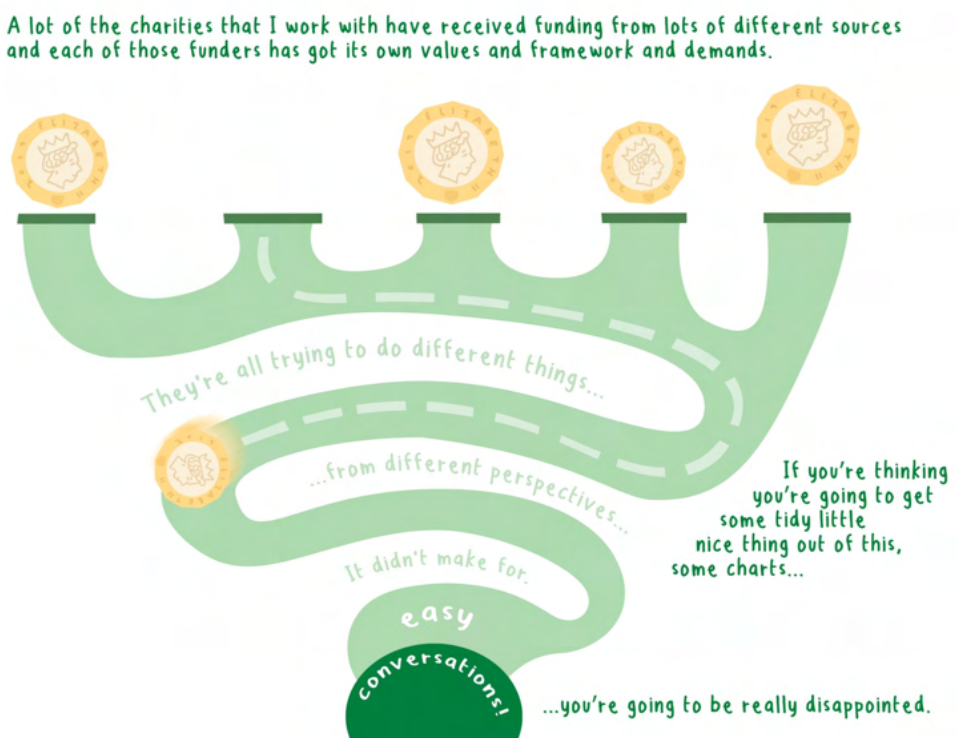
Getting the information that is
coming from the staff,
the artists, the relatives,
the good information,
and the reports,
into care plans.

it's the awareness raising bit.

The EVALUATOR'S problem

There's always this desire with dementia, because it's one of those incurables, is this going to fix it? Is it going to make their cognition better? There's a desire to have a long-term change and impact.

Having a good time this morning is one thing but does that impact people's behaviour longer term? Particularly people with dementia, is that the right question to be asking? because the here and now is so important. If we can make the here and now really enjoyable and exciting and colourful, is that enough?



The final part of the graphic narrative presents an account of the ways in which those I interviewed had overcome the challenges. These images can be found in the final chapter of this thesis (section 7.2.6).

5.6 Summary

This chapter has described the findings of analysis of stakeholder interviews. These findings show that these stakeholders experienced tensions, difficulties, and divisions as a result of value preferences and practices. They struggled to establish meaning in the context of arts activity involving people with dementia. And, when aiming for methodological 'rigour' or legitimacy, both of these challenges often came into sharp focus.

In Chapter 5, I reported findings from a series of interviews exploring the experiences of stakeholders involved in evaluation. Participants described tensions and divisions resulting from valuing practices and preferences as well as difficulties in making meaning in complex and difficult situations, and when aiming for rigour in their evaluations. In Chapter 6, these findings will be discussed and situated in the context of key concepts relevant to current research, practice, and policy in the UK. In the final and concluding chapter I will build on this discussion to provide some recommendations for future practice and research.

6 DISCUSSION: SEEING THE DILEMMA

6.1 Taking stock: ‘the way the world works’

The question that has driven this study is: What is it about arts-based activities for dementia that causes evaluators and researchers problems when describing, explaining, justifying, and implementing the methods they use to evaluate them?

In seeking to answer this question, I have consistently attempted to understand knowledge as both a condition and a continuing outcome of human agency, existing in constant interaction with the structures that surround it (Sayer, 1992). In line with the study’s theoretical underpinnings, I think of the production of knowledge as a social and relational activity (Law and Urry, 2004). To understand why evaluation might be challenging, it was therefore natural for me to explore the way in which people involved draw upon their cultural resources, experience, and backgrounds to apply various kinds of knowledge, skills and judgements to other knowledge ‘raw materials’ or data, thus re-presenting, producing or transforming knowledge in the process. In other words, their involvement in evaluation.

In presenting the findings of this exploration, I showed how these processes involved my study participants in applying and questioning their understandings of what is valuable and what might be meaningful about people with dementia engaging in arts activity, as well as about the methods and models that are used to evaluate that engagement. Close attention to their valuing and meaning-making practices revealed their anxieties and highlighted divides. In addressing these, evaluators were seen making judgements and compromises, adopting roles, and building or attempting to build relationships – and failing to do so – in the complex contexts in which people with dementia take part in the arts. They consistently struggled in bringing what they felt should be greater methodological rigour to the evaluations they delivered.

At an earlier stage, reviewing the literature around arts and dementia to identify and understand the idea of methodological challenge, I found that while researchers reported multiple challenges relating to study design, recruitment, and data collection and analysis, there were other key conceptual challenges woven through the fabric of their experience. These were the ideas of value, meaning, ethics, context, and use.

It is important to understand that arts and dementia evaluation may be delivered in a variety of ways: as academic evaluation research studies, sometimes commissioned and paid for by arts or third sector funding partners, sometimes by academic research bodies; as evaluation consultancy conducted by external independent evaluation experts; as part of a reporting system within an arts organisation or between an arts organisation and its funders, or; as reflective practice, both formal and informal, often embedded within artists' own ways of working. In most of these delivery forms, except (usually) the last, evaluation is a largely technical or 'top-down' process.

My findings support previous observation that evaluation activity in the arts may be valued more for its potential for advocacy or demonstrating accountability than for its capacity to contribute to knowledge or to improve practice (Merli, 2002; Belfiore and Bennett, 2007; Davies and Heath, 2013). A lot of it goes on. Every time an arts organisation delivers a funded project, an evaluation report of some kind is created and sent to the funders, where the information in it is – presumably – processed in some way. To illustrate: it was reported that a research manager at a specialist arts funding body received upward of 200 project reports each year, each of which was read and then coded in NVivo and the results stored as part of their ‘learning bank’ (Melville, 2017, pp. 55–56). It is worth noting that in this example however, there was no reported indication of how the funding agency proceeded to actually *learn* from this process of aggregative analysis.

Despite disagreement among evaluation theorists, including the observation that evaluation generally fails to live up to most of the promises it makes (Pawson and Tilley, 1997), it is clear that evaluation practice in arts and dementia remains strongly influenced by the idea that it is possible to obtain objective, ‘value-free’ knowledge, predominantly through the application of positivist-informed, often quantitative methods (Sanderson, 2000).

I have shown how beliefs about the value of particular kinds of methodological approach sometimes framed the decisions stakeholders made about the aims and objectives of evaluation: “… really what they wanted was proof of what they’d observed. […] They said, we want a randomised controlled trial.” (Harriet, researcher). Although this is the subject of considerable debate in the academic world, the legitimacy of particular methods, as well as ideas about what they might achieve were often accepted by those I interviewed as self-evident – if sometimes annoying – facts of life. One described this (with some frustration) as ‘the way the world works’ (Charlotte, researcher).

This thesis has also aimed to situate arts and dementia evaluation activity squarely in the context of the wider structures that inform its commissioning and delivery. My exploration of stakeholders’ evaluative activities should be understood in relation to a background of policy affecting the arts and culture, and dementia research. It also serves to highlight how the kinds of challenge identified have implications for the ethical conduct of evaluation; it points to important issues for the involvement of people with dementia in evaluation research more widely.

The product of evaluation is sometimes thought of as *evidence*, a term with particular currency in the context of health research and policy, where evidence-based practice is the norm. In the processes in which evaluation stakeholders are involved on the ground, the practicalities, politics, and ethics of evidence production are not easily untangled (Morse, 2006). This discussion chapter aims to start the process of loosening the knots.

6.2 Values and valuing

The findings of this research have shown evaluation stakeholders – my interviewees – routinely involved in addressing, or attempting to address, enquiries in which values play important roles. The methodologies they have used are also fundamentally value-laden, although they may not always recognise this.

Values have been described as psychological heuristics: in telling us about how the world *ought to be*, they can be useful in providing quick answers to complex problems. Sometimes we overlay them with emotions (Kelly *et al.*, 2015), and this can make them traps for the unwary if they are not signposted or acknowledged (Kahneman, 2011). We can term values *epistemic* if they contribute towards the goal of attaining knowledge – where they involve judgements and decisions about what is true, or simple, or can be explained, for example. Other kinds of values might be termed *non-epistemic* (Elliott, 2017) or perhaps *social* (National Institute for Health and Clinical Excellence, 2008) and these might involve judgements or decisions such as those about whether something can be said to be good, or just, or maybe beautiful.

The distinction between the two has practical application. For example, NICE distinguished between these two kinds of judgements to support its work in developing guidance for practice which might involve 'social value judgements':

> Scientific value judgements are about interpreting the quality and significance of the evidence available; social value judgements relate to society rather than science.
>
> (National Institute for Health and Clinical Excellence, 2008, p. 4)

The authors of this guidance go on to provide examples of the difference, focusing on situations that involve clinical judgements, including those with implications for discrimination or inequality.

In a more abstract way, philosophers of science and sociologists have been arguing over the status of value judgements in science for a century or more (Keuth, 2015). While rehearsing these arguments is beyond the scope of this thesis, my analysis of stakeholder experience shows what a struggle it can be to keep the two kinds of judgement separate in the way that NICE might suggest from the quote above to be possible.

In presenting my findings I talked about *valuing practices*, by which I mean any activity involving an individual in actively applying an idea of value to some practical end, such as making a judgement based upon it. In the evaluation context this would include making decisions about what question to ask, to whom, and how to apply or disseminate the answers. It also involves decisions about whether particular methods, approaches, and the subject or aims of evaluation seem ethically good or fair.

Differing perceptions about what kinds of knowledge are valuable created difficulties for evaluators working in the cross-sectoral and multi-disciplinary partnerships that characterise the field of arts and dementia. Understandings of why it might be valuable to participate in arts activity were not always shared by project partners. The funding and resourcing structures of evaluation projects and the contexts in which the arts are commissioned, as well as that of dementia treatment and care, also reflect valuing practices and must be recognised as contributing to creating a challenging environment for evaluation. I have suggested that arts activities for people with dementia prove hard to evaluate in part because of the cultures, structures, and valuing practices prevalent in the field. In a later part of this chapter I will seek to demonstrate how attempts to reconcile epistemic and non-epistemic values result in evaluators making methodological compromises and trade-offs (section 6.2.1).

The different epistemic values evident in arts and dementia evaluation practice continue to reflect the divide between the sciences and the humanities. Put crudely, the criteria we use in judging the worth of knowledge gained through research or evaluation are likely to differ, depending on the discipline in which we have been trained or in which we practice. Examples of this were described by those I interviewed. For academics, and in relation to methods for research or evaluation, this divide is sometimes characterised further as that between methodologies embracing a positivist worldview and their anti-positivist or interpretivist alternatives, although in reality things are rarely that clear cut.

As the experiences of my interviewees show, the practices and preferences associated with world-views on either side are not confined to academic studies where their pros and cons are generally well-understood and well-signposted in reporting. Instead, they seep, perhaps unacknowledged, across boundaries and into the language and valuing practices of people from the health, arts, or care sectors. They rise to the fore during evaluation activity.

Those who are not evaluators or researchers are likely to have only a ‘fuzzy’ idea of what the difference between evaluation and research might be (Wanzer, 2019). As an additional complication, the boundary between research and evaluation disciplines has been described as ‘unhelpfully porous’ (Raw and Robson, 2017). Language and approaches from the research world slip into evaluation discourse away from the academic context. Sometimes they take on the character of talismans or rhetorical devices – used to indicate the presence of epistemic values such as rigour, objectivity, and credibility. They do not, of course, necessarily ensure any of these things and their epistemological origins may not be clear or understood by those who use them.

Interviewees described misunderstandings around the purposes and strengths of different evaluation methods, approaches, and measures. They provided examples of their resulting and occasionally inappropriate application, including (but not limited to) qualitative study employed to do the work of quantitative. As an example, both in conversation, and in written evaluation reports, there was evident confusion around terminology such as *impact* and *outcome*, although this perhaps reflects a more general lack of definition around these terms (Selwood, 2002). Artists or arts managers also used words such as *evidence*, *objective*, *significant,* or *rigorous* in ways that differed from their academic use (where it should be noted again that they are also contested). And finally, the word *evaluation* was itself differently defined, with one artist – for example – talking about how she used stories to evaluate and communicate her practice to others, another talking about having an evaluative ‘practice of enquiry’ around her work.

Negative views about evaluation and its worth were prevalent amongst arts practitioners and this had affected relationships and communication during evaluations. Similar tensions in evaluator-artist relationships have been identified elsewhere, where it has been suggested that the logics that might explain these may have their root in the way in which artists relate to the institutions that pay them (Melville, 2017).

However, despite expressed anxieties and concerns from arts practitioners in particular, my research suggests the epistemic divide in arts and dementia may be less starkly drawn and could be more productive than anticipated. Few of those I interviewed were set or comfortable in any single methodological position, but practitioners and evaluators sometimes crossed, acknowledged, and even worked with the divide. So, for example, an experienced dementia researcher with a background in positivist research acknowledged:

> I think there's something about having been trained in a positivist framework or whatever. […] most of the world probably doesn't work in that way. And particularly creative practice, I guess, doesn't. By definition'.
>
> (Mary, dementia researcher).

An arts manager (Philippa) described what she saw as a difference between approaches driven by acts of 'intervention' and those informed by 'interaction', and described attempting to work with both within the evaluation frameworks and practices of her organisation. And an artist (Deborah) described her attempts to find meaningful ways of measuring the effects of her work, drawing on principles of reflective practice and her understandings of wellbeing measurement in assessing how positive health and wellbeing was observably manifested for the individuals with dementia taking part in the sessions she delivered. These examples indicate that those involved at the sharp end of evaluation practice may perhaps be thinking beyond measurement of impact and towards understanding of the processes and mechanisms that enable it.

6.2.1 Valuing evaluation

In the case of arts and dementia projects, important non-epistemic values will include the principles held by those involved that relate to the arts and culture and how and why it might be important to evaluate our participation in it. They will also include those that relate to the consideration, care, and treatment of people living with dementia. It is useful to recognise that any and all the value preferences adopted as part of evaluation criteria will have an effect on the methodologies we use and the interpretation of results (Kelly *et al.*, 2015). However, the boundary between epistemic and non-epistemic may not always be clear, and there is justifiable argument about whether the roles values play in research or evaluation are legitimate or not (Elliott and McKaughan, 2014; Elliott, 2017).

In line with previous research (Goulding, 2014; Daykin et al., 2017), my interviewees confirmed that arts and dementia evaluation is under-funded and under-resourced. Most evaluation research studies (as opposed to the more prevalent commissioned summative or service evaluations of particular projects) are small-scale: seventeen of the studies included in my literature review described themselves as pilots or exploratory. In dementia research it has been reported that exploratory studies ‘that indicate positive outcomes of nonpharmacological interventions are often underfunded and subsequently discounted as not rigorous enough’ (Zeisel *et al.*, 2016, p. 2). This may be a symptom of a wider problem still, since while exploratory trials have been recommended by the MRC as part of a larger process, there is reportedly limited guidance available on how they should be conducted, or when they might be useful (Hallingberg *et al.*, 2018; Moore *et al.*, 2018). While guidance on the evaluation of health-related interventions recognises cost as a valid consideration for choice of method (Craig *et al.*, 2008), authors of research studies into the arts and dementia routinely recommend more ambitious and more rigorous experimental studies. The cost of conducting these is a concern, and well out of reach for most.

More worrying still, global investment in research into health and social care innovation that will improve the quality of life of people with dementia has been described as ‘woefully’ absent (Pickett and Brayne, 2019). In the UK, in contrast to spending on health, public spending on social care has fallen consistently since 2009-10, despite increasing cost and demand pressures (Charlesworth *et al.*, 2018). My interview with a dementia specialist sitting on a regional Clinical Commissioning Group confirmed that budgets and financially incentivised targets were of primary concern for her and her colleagues, with these including a recent directive to increase dementia diagnosis rates. Evaluators and researchers involved in projects operating at the intersection between the systems for arts funding and non-pharmacological interventions face challenges at systemic levels. These include the likelihood that exploratory or feasibility evaluation studies and their results may be undervalued and their insights not followed up, and that funding is unlikely to be available to conduct larger scale trials.

Arts and dementia projects and programmes described by those I interviewed were also frequently time-limited and dependent for funding on a patchwork of sources. Evaluators were competing for capacity and resource alongside implementation of the arts activity itself and so evaluations were often light touch, with evaluators making do with what was available to them: training others or relying on untrained colleagues (often artists) to collect data, for example. This led to difficulties at the point of analysis, when insufficient evaluator time was available to make use of large quantities of data gathered, or when data were discovered to be of poor quality. If an evaluation involved an ethics review or even simply a consent process, artists and care staff might be drafted to gather consents, meaning that evaluators could not always be certain that correct processes had been followed. Evaluators also sometimes limited the number of points at which data were collected, potentially ruling out the use of some methods which show promise for predicting changes in wellbeing for participants, such as regular experience sampling (Holt, 2018), observations that might illuminate details of process and change over the course of a long-running programme (Vella-Burrows and Wilson, 2016), or intensive video observation that can reveal micro-patterns of interaction and communication (Clare *et al.*, 2019).

The health and care settings in which many activities took place were themselves operating under financial constraints and some were described as ‘business-driven’. Interviewees suggested that decisions made within and for them by commissioning systems had key drivers which largely related to efficiency, effectiveness, and value for money. As an indication of how the values of a context in which evaluation or activity is conducted may shape an evaluation, several arts project managers recommended these drivers should be reflected in the aims and objectives for evaluation in care settings, referencing for example, the standards of the Care Quality Commission.

Many evaluation projects described by interviewees had taken place in residential care settings. In these, financial instability and organisational uncertainty were described as common. Evaluators reported patterns of high staff turnover, shortages of qualified staff, care workers working long and unpredictable shifts, and instances of poor literacy in staff whose first languages might not be English. They reported situations in which care homes had closed part-way through a project or where care partners had pulled out without explanation or notice before a project began. Care work was often described as under-valued. These factors were reported to have swayed evaluators' choices of methods and study design and had sometimes threatened to or actually compromised data collection.

It was also noted that staff involvement and enthusiasm could be key factors in the success of arts projects in care settings. Several evaluators reported low levels of engagement in evaluation, especially from management, and particularly where commitment of time or money was required. The peripheral role that arts and creative practice is seen as playing within healthcare, coupled with its delivery by what interviewees' experience suggests is an atomised and largely unprofessionalised workforce, was felt to have led to the contributions of arts practitioners lacking visibility.

6.2.2 Valuing the arts and questions of quality

Differing value practices were also identified among interviewees in relation to arts participation itself. All the arts and dementia projects described were devised and managed by arts organisations. Most were funded by arts, charitable funders, or local authorities. All, however, were designed to provide health and/or wellbeing benefits, and most were delivered in health or care settings.

Artists interviewed were purposively selected in line with the focus of the thesis, with one of the selection criteria being that they defined their practice in contrast to therapy. As such, most placed an emphasis on the processes of participatory art-making rather than on delivering specific health and wellbeing goals for individuals or groups:

> I’m a dance artist working in participatory dance settings as opposed to a therapist trying to fix people.
>
> (Deborah, arts practitioner)

> […] making music together, that everybody feels that they have been heard within that, and if that has a therapeutic benefit for anybody then that’s great […] that’s not my intention and it’s not my expertise.
>
> (Justin, arts practitioner)

While arts therapists may have standards, frameworks, and training or education already in place, there are no consistent or agreed standards for arts practitioners, and little agreement about how to ensure the quality of their practice. It is perhaps not surprising then that questions of quality, and how to assess it, arose during interviews.

It has been argued that quality in participatory arts is particularly challenging to evaluate because of a requirement that it address the varying value frameworks of those involved, alongside what is described as the inherent subjectivity of artistic experience (Matarasso, 2000, 2013). Definitions of quality are of importance because unpicking them essentially reveals what different people perceive to be of value about a service or activity.

Among artists and arts managers interviewed, informal judgements of quality are perhaps most easily traced to the world of participatory or socially-engaged arts practice. Here process is emphasised, with the *art* residing in the action and interaction between artist and participant, rather than within a finished artwork (Matarasso, 2013). When this process involves people living with dementia, certain value aspects of it may receive particular emphasis, perhaps those that relate to principles of inclusion, equality and fairness, and respect for the autonomy of individuals.

Interviewees ascribed importance to activities in which experienced practitioners interacted in ways that demonstrated understanding and sensitivity to the needs and requirements of people with dementia as well as an ability to establish meaningful connections with individuals. Practice was prized that was non-directive and flexible, took its lead from participants and incorporated an openness to experimentation and attentiveness to ‘in the moment’ interactions. This included work in which the varying skills and strengths offered by different artforms were combined. Some practitioners described a move towards what they termed ‘co-creative’ practice, in which the distinction between participant and artist is blurred, and the contributions both make within an arts activity more equally recognised.

Other markers of quality were described. Some may have their origin within the established art-world and its notions of aesthetic value or excellence, where intrinsic criteria and the final products and outputs of arts practice are privileged: for example, a book of words beautifully printed and designed, a public exhibition of artwork, or a performance that attracts public audience. There were also the markers that relate to a familiar set of instrumental factors such as wellbeing or quality of life, or alleviation of the behavioural and psychological symptoms of dementia. Of these, positive improvement in mood or wellbeing were the outcomes most evaluators interviewed had sought to identify.

Interviewees understood that outputs might be measured through monitoring, and that a variety of different measures were available to assess whether arts practice had achieved its desired instrumental outcomes. However, markers of quality relating to successful engagement in processes of participatory art were considered particularly hard to pin down – perhaps because this area of quality is more closely tied to judgements made on moral or ethical principles. Interviewees felt this kind of definition of quality might not be well understood outside the arts world, including among health or care staff.

The concepts of *quality*, *impact* and *value* are inter-related, and the terms were sometimes used synonymously by artists I interviewed. While there is no universally accepted definition of value or quality in arts fields, for participatory arts, through the ubiquity of the concept of impact in relation to policy and funding, quality has sometimes been linked to cost and effectiveness. Methods of econometric analysis such as social return on investment have been used to establish the significance of the arts and to make the case for its continued subsidy (Oman and Taylor, 2018).

In contrast, *quality* is sometimes used as a handy, if potentially misleading, proxy for *good*, reflecting an emphasis that has at times been placed on ‘excellence’ by national funding bodies. Excellence is a concept ACE has rather unhelpfully admitted is both relational (cannot be separated from the people that value it) and that will mean different things for different disciplines, organisations and ways of creating and presenting work (Arts Council England, 2013). Discussions of quality reveal stakeholders in arts and dementia evaluation struggling to balance perspectival factors in conceptualising ideas of quality arts practice with the need to advocate for particular (better than others) iterations of an activity, or for the artist or organisations delivering them. There are obvious implications for the reporting of negative or anomalous experiences. It is disappointing, but perhaps not surprising that the ethical implications of arts practice involving people with dementia that does not meet standards of quality have received very little evaluation or research attention. Concerns around this issue were, however, an identifiable theme arising from my interviews for this study.

As with the arts there is also no universal quality standard for healthcare (The Health Foundation, 2013) or dementia care, although several bodies use standards that apply broadly similar categories. For example, the US Institute of Medicine (IoM) describes quality as: ‘the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge’ (Institute of Medicine, 2001, p. 44). Within this definition, it identifies six dimensions – safety, effectiveness, patient-centeredness, timeliness, efficiency, and equity. In 2012, the NHS Commissioning Board identified the six C’s of care, compassion, competence, communication, courage, and commitment as central components against which the quality of practice of nursing, midwifery, and care staff might be assessed (Cummings and Bennett, 2012). The CQC regulates care settings in the UK and upholds fundamental standards that include person-centred care, dignity, and respect as well as physical, safety, and organisational commitments. For dementia care, the Care Fit for VIPS framework also breaks the concept of person-centred care into component parts, this time with the aim of ‘making services better’. The acronym reflects the suggestion that quality person-centred care should include attention to Values, Individual needs, the Perspective of the person, and supportive Social psychology (Brooker and Latham, 2016).

A discussion of *quality* in the light of such definitions provides a reminder of alternative angles from which the challenge of understanding the value of participation in the arts might be approached. It is perhaps useful to consider the

Health Foundation's suggestion that measuring for quality improvement differs from that conducted for research or evaluation, in that it involves an approach in which learning develops and informs the process and in which it is accepted that data may need only to be 'good enough' rather than perfect (The Health Foundation, 2013).

When we explore ideas around quality, we are beginning to unpack some of the processes and mechanisms involved in arts and dementia practice. This also means that we are doing work to identify what the people involved feel to be important about it. When considered in the light of the preceding discussion of values, we might also want to consider the implications of both quality and its improvement being thought of as temporary, local, and dynamic.

6.2.3 Some prevailing world-views

As has previously been discussed, in healthcare the prevailing paradigm for knowledge production and utilisation remains that of evidence-based practice. Here the MRC guidance on developing and evaluating complex interventions (Craig *et al.*, 2008), underpinned by the experimental method, has been formative in shaping evaluation research. It has also informed development of public health evaluation frameworks both in and outside the health sciences, including those for evaluation of arts for health and wellbeing (Fancourt and Joss, 2015; Daykin and Joss, 2016; Fancourt, 2017). The MRC guidance recognises the organisational and logistical difficulties of applying its methods to service or policy change, and the difficulties of linking interventions with outcomes, two qualifications that might make arts and dementia evaluators warm towards it. However, as Fancourt and Joss (2015) have noted, an absence of specific reference to social or arts-based interventions in the MRC guidance makes it difficult for evaluators to know how to apply it in practice in these contexts.

The research processes around evidence-based practice might appear to privilege particular kinds of evidence and methods over others. However, this does not necessarily represent its true ends and aims; early adherents of EBM recommended that, rather than slavish adherence to rules or 'cookbook' approaches, it should entail 'tracking down the best external evidence to answer our clinical questions’ without automatic assumptions about the methods this quest might require (Sackett *et al.*, 1996, p. 72). The MRC guidance favours inclusion of a range of methods. It recommends fitting the method to the question, and provides the observation that ‘best available’ methods may yield useful results even if they are not ‘theoretically optimum’ (Craig *et al.*, 2008, p. 9). In opposition to hierarchies, a ‘horses for courses’ typological approach to evidence production and review has been recommended, within which methodological appropriateness rather than study design was emphasised (Petticrew and Roberts, 2003). The idea that typological rather than hierarchical assessment might be appropriate was introduced as a part of the literature review for this research (see section 2.4). Such an approach means that attention can be focused on the fit between the research question and study design, rather than on assessing the relative methodological rigour of evidence presented.

In the wider field of health research there is a move towards a tempering of EBM through approaches that more fully recognise the contributions and needs of practitioners, managers, and of lay people in improving practice (Greenhalgh, 2017). This has been accompanied by discussion around exactly how the differing kinds of knowledge they will provide should be integrated within guidance (Wieringa, Dreesens, *et al.*, 2018).

In the framework developed by Fancourt and Joss (2015) for evaluating arts and health programmes the emphasis is still placed on increasing technical and methodological rigour rather than providing guidance on how evaluation approaches might contribute to understanding or improving practice, or reflect the values of participants and practitioners. However, the authors do try to ‘create space’ for and to ‘incorporate’ other epistemologies and methodological paradigms including – for example – Participatory Action Research.

Theory-based approaches, including realist evaluation, have found favour in assessing the social impacts of the arts – perhaps fuelled in the UK by the influence of the ‘What Works?’ agenda (Galloway, 2009; All Party Parliamentary Group on Arts Health and Wellbeing, 2017). Methodological debates simmer, nevertheless, over the question of whether randomised controlled trials are or are not compatible with realist approaches and – if they are – what their limitations might be when applied to complex public health interventions (Van Belle *et al.*, 2016).

In the field of cultural policy it has been proposed that evidence-based discourse does not provide an appropriate paradigm through which to assess the value of the arts, either for individuals or for the public sphere (Belfiore and Bennett, 2007). As an alternative, it has been suggested that critical and humanities-based studies will produce insights that can speak usefully to a desire amongst policymakers for enquiry that enlightens, educates, and can help to advocate for change (Belfiore and Bennett, 2010). Others have argued that these studies can provide illumination of process and underpinning theory in relation to individual or social impact (Galloway, 2009; DeNora and Ansdell, 2014).

These alternative viewpoints are not the norm for work in the fields of the arts that seeks to make the case for their importance to society, however. Cultural policy has, since the 1980s, focused largely on pursuing instrumental arguments to justify government spending on arts and culture. In such a framework, arguments for the social value of the arts have been closely linked to the economic benefits they are thought able to provide. Arts and culture have been presented as net contributors to the economy with roles to play in – for example – tackling social exclusion or regeneration (Myerscough, 1989), or – more recently – health and wellbeing (Fujiwara, Kudrna and Dolan, 2014). For a number of years the emphasis for evaluation in the UK cultural sector has been on providing *auditable* evidence of its impact in forms intended to support its judicious application in current politically relevant areas (Belfiore, 2004). Whether this may come at the expense of arguments for more intrinsic benefits has been the subject of intense debate, including suggestions that the focus renders the arts merely a 'tool of government policy' (Holden, 2004).

If instrumental arguments for the value of the arts are made on the basis of their ability to achieve social and economic objectives, for which the outcomes could be achieved by other means, then it doesn't seem unreasonable that analysis of their utility should bear comparison with those alternative means. In such situations, as was noted by one academic researcher I interviewed, if the cost benefits of the arts are perceived to be minimal, even a positive evaluation of impact is likely to carry 'less weight'.

To this end, economic or cost-benefit analysis focused on the arts activity has been recommended as helpful for the field (National Endowment for the Arts 2013; Camic et al. 2013; Fraser et al. 2014; Zeilig et al. 2014). However, as my literature review demonstrated, it is still rare for it to be conducted for the arts and dementia. While there was one economic evaluation in the grey literature (Fujiwara and Lawton, 2015), there were no published studies. This is perhaps because when economic impact studies have been conducted in other areas of arts and culture, their quality has been criticised (Jensen, 2014; Crossick and Kaszynska, 2016). The econometric models used to create them have also been described as 'impenetrable' to those in the arts sector, requiring the involvement of independent expert consultants for their application (Oman and Taylor, 2018).

The problem of how we construct, operationalise, and then measure the value of the arts and culture more generally, including its social impact, has been the subject of discussion and debate in academic and policy circles since the later years of the 20th century (Melville, 2017). In the 2000s, the term 'cultural value' came into currency, along with inconclusive debates about how it might be possible to value both intrinsic and instrumental values and whether we should try. An influential ESRC/AHRC-funded report for the Department for Culture, Media and Sport (DCMS) outlined the valuation techniques that the cultural sector might apply to its work, using HM Treasury's *Green Book* on policy appraisal and evaluation (O'Brien, 2010). In 2016, the final report of the AHRC Cultural Value Project called for a more loosely defined framework for the arts and culture, 'within which sits a variety of components of value, each of them to be evaluated by appropriate and often different qualitative and quantitative methodologies [of which] the fundamental criteria must be appropriateness to the subject and to the analysis, and robustness in how evidence is gathered and deployed' (Crossick and Kaszynska, 2016, p. 24).

In 2019, the UK's first 'Centre for Cultural Value' was instituted at the University of Leeds. The core questions it seeks to answer can be found on its website, where epistemic and non-epistemic value issues are placed firmly side by side:

> Why do arts and culture matter? What difference do they make to people's lives? And how can we evidence this?
>
> (University of Leeds, no date)

Prevailing winds lend distinctive shape to trees that grow in their paths. Guidance frameworks and centralised strategies generally emerge from within existing and contingent structures and hierarchies of evidence (Greenhalgh, Howick and Maskrey, 2014). However, people's experiences of the arts are so complex, paradoxical, and varied, depending upon who is doing the experiencing, that any single evaluation tool or framework to measure their value is likely to prove a disappointment (Walmsley, 2018). There is a danger that the application and production of such tools may crowd out alternative or complementary processes or fail to recognise the contribution of local, adaptive knowledge, thus leaving less space for individual perspectives such as those originating in the experiences of arts practitioners and those who use commissioned services. It has similarly been noted that particular focuses for evidence-based policy-making, such as the attention currently paid to subjective well-being, may tend to reproduce values and hierarchies rather than producing new evidence, with the result that certain forms of cultural practice are viewed as having higher value than others. This can lead to a recursive relationship between research, practice, and policy: 'the more an artform is researched, the stronger the arguments for funding for further research into its contribution to well-being, thus supporting public subsidy of that artform, its further research, and its increased visibility in cultural policy-related discourse' (Oman and Taylor, 2018, p. 226).

6.2.4 Advocacy

Commissioning structures which link funding with the findings of evaluation have – at least historically – been connected to the production of a 'proliferation of methodologically unsound impact studies' (Belfiore and Bennett, 2007, p. 137). It has been argued that independent consultants and academics alike have been encouraged to focus on providing evidence of impact at the expense of exploration of theory, process, and mechanisms. That this may be the case for those in the arts and dementia field is supported by those I interviewed, who sometimes described their role in evaluation as that of a tool used to achieve a particular end, or spoke of answering the questions about efficacy they had been told to ask, rather than exploring underlying issues. In an emerging field of enquiry historically driven by the arts sector, it still appears difficult for evaluation research partners to establish an agenda separate from the needs of arts partners wishing to make the case for public or private subsidy. It has also been argued that the process has left arts organisations and artists feeling 'besieged by inappropriate criteria' (Raw and Robson, 2017). Again, this is supported by the responses of arts managers and practitioners interviewed who expressed anxieties about the aims, process, and resulting use of evaluations in which they had been involved.

Interviewees reported feeling dependent on the results of evaluation and research to help them advocate for future work. As well as encouraging a focus on outcomes and impact, this had other methodological consequences, including an observable tendency for evaluation to be reactive, and to focus on (positive) outcomes or to explore semantic or surface level meanings in evaluation data.

Additionally, a reliance on short-term funding from multiple sources may be leaving arts organisations without a coherent approach to evaluation and to delivery of their work. While some stakeholders argued that funding instability had not compromised the core values of their organisations, it was reported to have had an effect on evaluators' ability and arts organisations' capacity to successfully negotiate, explore, and sustain extended engagements between health and care staff, artists, and participants with dementia. Stakeholders also reported difficulties in accounting for the extended role of arts and creative activities in interaction within the systems in which they were delivered, and in understanding their longer-term effects.

This discussion has explored the ways in which the challenging structural relationships between the value preferences of stakeholders and the methodologies used for evaluation leave arts and dementia researchers and evaluators in a perplexing place. They are required to carve out an agenda for evaluation research that accounts for cost, achieves some kind of visibility and serves practical purpose, but which does not follow well-worn impact evaluation tracks that can lead towards advocacy and reduced credibility for their results.

6.3 Finding or making meaning

The second key concept explored in the findings of this study involves the difficulties evaluators experienced in applying existing measures and tools – even those designed for people living with dementia, in the face of a condition which fragments the channels of communication between evaluator and evaluation subject. For many of those I interviewed, the evaluation of arts and dementia activities presented a perfect storm of a puzzle: how to make sense out of what some termed the *intangible* nature of experience and participation in arts activity, given that the person experiencing and participating is likely to have difficulties with cognition and memory. And although *meaning-making* is an activity generally associated with qualitative evaluation and research, it can be equally a problem for evaluators using quantitative methods, as some of the studies described in the literature review for this project showed.

Interviewed stakeholders were enthusiastic about creative and arts-based methods although there was little evidence that they were being used widely and there were concerns about how data elicited through them might be interpreted and integrated within a wider evidence base. They raised questions around the effectiveness of existing methods and measures for capturing the temporal aspects of arts engagement for people living with dementia. Phenomenological effects, relating to the experience of arts engagement were often termed intangible and the difficulty some interviewees had in describing them reflects a wider debate within the arts world about how and whether such effects can or should ever be subject to evaluative judgement. The gathering of contextual data (socio-demographic, biographical, clinical etc.) remains practically challenging for evaluators, although stakeholders suggested that an understanding of the context surrounding the core data they were collecting was necessary to support and sometimes enable its interpretation.

6.3.1 Fragmented communication: involving people with dementia

Some research practices have been described as ‘gagging’ for people with dementia (Brooks, Savitch and Gridley, 2016). This is largely because of assumptions that there will be difficulties with consent and capacity and because of the persistence of a belief that people with dementia are not able to communicate in a form that researchers will find useful. While the ethical participation of people with dementia in research is recommended (Alzheimer Europe, 2011), and their collaboration increasingly sought, yet the range of methods regularly employed to enable this involvement is still limited (Bartlett, 2012; Phillipson and Hammond, 2018).

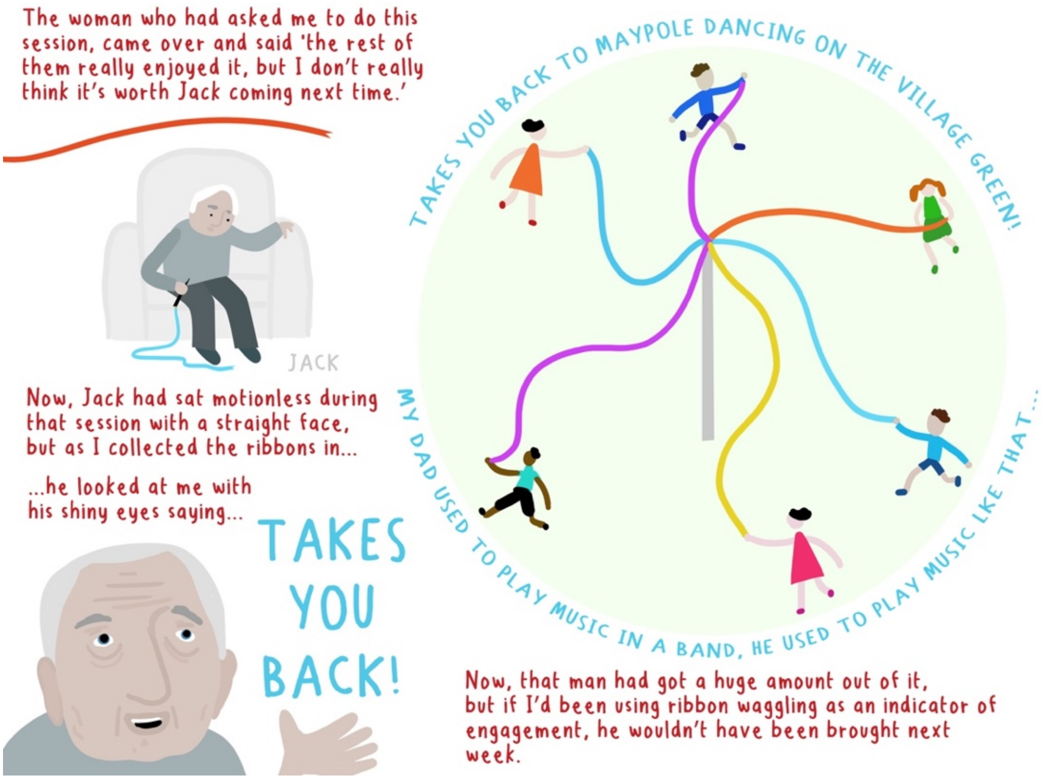
Evaluators I interviewed had used the standard research methods of survey and interview or focus groups, but often found them unsatisfactory. Many standard self-report measures are not designed for respondents experiencing difficulties with cognition, memory, and communication. Even those measures that are designed for people with dementia will prove difficult when dementia is more advanced or if a particular form of dementia does not allow communication in the way the measure demands. To elicit meaningful results from survey instruments or verbal interviews, evaluators will need to be skilled in working with people living with dementia and should be prepared to devote considerable time and patience to the endeavour (Cridland *et al.*, 2016). Time and resource were not readily available to arts and dementia evaluators taking part in my study.

The barriers evaluators experienced to using standard research formats such as interviews and surveys sometimes led to a reliance on proxy or informant accounts, with caregivers (both professional and family) being asked to answer evaluation questions for the people in their care. Some of the problems of this were explored earlier when discussing findings of the literature review (section 2.5.3); it is well understood that proxy accounts may differ from those that are articulated by persons with dementia themselves (Murphy *et al.*, 2014). Although the responses of informants who are carers were sometimes described as insightful, my interviewees' experiences revealed how evaluation methods can sometimes play a role in eliding the identity of a person living with dementia with that of the person caring for them.

Evaluator interviewees also found observation, a standard alternative to oral or written questioning, problematic. This was particularly the case when resources were not available to conduct regular observations or to analyse data, or when skills and training were not in place. There are concerns that some structured observational measures may either be too specific or too generic for use in the context of an arts activity with people with dementia; in both cases perhaps failing to adequately capture the constructs they were designed to assess.

With structured observations evaluators had to accept that things might be missed because they weren't part of the observation framework, or if they occurred outside the observational time window. This was something that artists in particular found hard to stomach, and they often provided stories to demonstrate the importance of moments that had not otherwise been made available for analysis. My work with an illustrator to create dissemination assets for this project neatly encapsulated one of these stories (see Figure 4 below). It also shows how an evaluation method can have real, as opposed to merely epistemological, impact – and suggests that this impact may have an ethical dimension.

Figure 4: Graphic narrative 'Takes you back'



Some of the difficulties of field observation might have been overcome through the use of video, since a camera makes fewer initial editorial decisions and analysts retain the option to slow, pause, and review film, making both micro and macro patterns of action and interaction visible. Film can also easily be taken back to participants or proxies to check the meanings a researcher has derived from it, or viewed by analysts representing different disciplines. However, using video with people with dementia and in dementia settings also requires skills and resources and can involve significant ethical challenges (Puurveen *et al.*, 2015).

While this study has shown evaluators applying some arts-based methods, primarily these were simple tools used to elicit data relating to participants' mood or feelings of wellbeing. Participants had – for example – been invited to create a visual representation of their mood, or to choose an image or colour swatch representing their feeling. Such methods were sometimes described as more acceptable to participants or felt to be in tune with the content of the arts activity itself. But, where evaluators had tailored, created, or adapted such tools (or other measures and surveys) to fit their individual projects, findings then had little validity beyond the context of the individual study. As has been observed for evaluation in the museum and galleries sector, evaluation results – even if they are publicly circulated, which is not always the case – may not easily find a place within a wider body of findings or recommendations (Davies and Heath, 2013).

One of the researchers I interviewed had used an ethnographic observation framework, but most evaluators had been conservative in their choice of methods. Approaches of the kind that might be said to hail from the 'edgelands' of qualitative enquiry for health research (Rapport, Wainwright and Elwyn, 2005), such as narrative-based or anthropologically-informed methodologies are rare in the field, as was demonstrated in the published studies for my literature review. More generally, this may reflect an uncertainty about the role that qualitative approaches ought to play within evaluation.

Innovation and modifications of standard methods and tools could offer valuable changes of perspective and an opportunity to rebalance traditional research positions of control and responsibility. However, their successful application requires a confident and supportive research and evaluation environment; flexible ethical processes, and time and energy to devote to building meaningful relationships with participants are also considered essential (Phillipson and Hammond, 2018). Arts organisations and evaluators are also likely to need reassurance that any method they choose will be acceptable to those who will read the report and to funders who might commission it.

Evaluators of arts and dementia projects interviewed both perceived and actually experienced significant barriers to their inclusion of people with dementia as direct informants in studies. In some evaluation studies this appeared to have motivated a turn of evaluation focus away from the person with dementia and towards the care context. Although perfectly valid and interesting as a research direction in itself, if coming at the expense of attention to the individual with dementia, the risk is that this will perpetuate the perception of a person with dementia as a subject of care rather than someone in possession of a continuing capacity to engage with all aspects of life – including the arts. Personhood may be eroded as a result. It also doesn't sit well with recent calls to recognise and make visible the rights of people with dementia to not be treated differently because of their condition, while adding to the evidence that they still are (All-Party Parliamentary Group on Dementia, 2019).

6.3.2 Measuring the intangible and the momentary

It was notable that stakeholders interviewed sometimes confidently referred to effects of the arts using short-hand, borrowed, or contested terms such as wellbeing, quality of life and 'in the moment' but they became noticeably less fluent when trying to describe what these actually meant in practice (section 5.3.2). In addition, there was a commonly held belief that some effects of the arts will only ever be momentarily observable. For some this could even extend to a belief that positive changes in some dimensions of health and life for people with dementia are simply not possible. This is despite research indicating efficacy of the arts in contributing to positive improvements in many areas, including – for example – specific functions such as episodic memory and verbal fluency (Eekelaar, Camic and Springham, 2012; Palisson *et al.*, 2015).

A lack of fluency in describing both the effects and the experience of taking part in arts activity, and a resort to terms such as intangibility and ideas of transience also reflect the arguments of the instrumental / intrinsic debate in cultural policy. An evidence review produced for ACE in 2014 urged its readers to 'always start with the intrinsic' asserting that arts and culture's inherent value could never be quantified, but that, nevertheless, organisations must continue to try to do just that in order to secure funding (Mowlah et al. 2014; Carnwath & Brown 2014). In the same vein, some artists told me that they couldn't comprehend of the possibility of quantifying experiences, although they accepted that the attempt had to be made. For example: 'you can't measure atmosphere, it's just a… it's a feeling thing isn't it? It's not… it's not… it's not concrete. So… so… I can see the dilemma there.' (Coral, arts practitioner). Few artists actively denied the value in approaches that enabled quantification of changes or impact; in Coral's words, they 'see the dilemma'.

Methods drawing on the reflective valuing practices of artists may provide an alternative to the standard research approaches and indicate the potential for an application of different methodological paradigms to the problem. Arts practitioners are often asked by evaluators to reflect on their experience of facilitating arts sessions, formally and informally. Sometimes they will be asked to provide written accounts or journals about the work they do, but evaluators struggle to understand how to use these often highly subjective accounts.

Reflective practice methods more generally have been recommended for their capacity to illuminate process and to contribute understanding of the organisational structures that may facilitate or create barriers to positive experiences of the arts for participants (Melville, 2017). It has been reported that qualitative evaluation practice incorporating collective reflection and individual journaling by artists, in combination with interviews and participant observation by an evaluator, can contribute to professional development and creation of positive relationships between artists, the delivery organisation and the evaluation team. It can also go some way towards satisfying funder's requirements for evidence and accountability (Raw and Robson, 2017).

In experimental studies contextual information may be gathered to help control or account for variables that can confound results. However, interviewees observed, many different kinds of contextual information can be helpful in making sense of the intangible and sometimes momentary effects observed. Understanding of the context surrounding an observed effect supported and sometimes enabled interpretation: details of a life history making sense of a participants' concentrated attention to the minute details of a task, for example. Reflection and reflective practice were commonly used by the artist interviewees to support such understandings (section 5.3.3).

Viewing the participants in arts activities as people living, making art, ageing or connecting with others, rather than as patients with physical and medical needs, might encourage evaluators to consider other ways in which to explore the value and meaning of their participation. Rather than a series of momentary engagements whose effects are measured using metrics that assess biological, psychological, or cognitive symptoms, an experience of arts activity might fruitfully be considered as one interacting element within a wider ecology, in which it then becomes possible to include other data points. Evaluators might then choose to explore the relational responses and interactions occurring within this ecology, and over time – moving beyond the momentary encounter. These might – for example – further elucidate the ways in which the arts can provide vehicles through which people create and experience communicative and multisensory connection with others (Clare *et al.*, 2019) or the physical world, or perhaps serve as places of haven (Garabedian and Kelly, 2018), escape, or fantasy.

In the caring professions, reflection has been shown to have the potential to move practitioners from 'knowing how' towards forms of 'practical wisdom' that can guide ethical action and interaction with patients (Kinsella, 2010). Releasing artists' reflective practice from the business of providing evidence of impact – a role for which it is not suited – may prove helpful; turning experiential learning into action for change in a way might also align with participatory action research methods (Baum, MacDougall and Smith, 2006). Attention to the mechanisms underpinning artists' reflection *in* or *on* action (Schön, 1983) could help reposition arts activity with people with dementia as an ethical and 'careful' practice in which the needs, feelings and perspectives of both participant and practitioner are recognised and their values balanced and accounted for (Tronto, 1998).

6.4 Making compromises and building relationships

In the first sections of this chapter, epistemic values were discussed in a general sense, in order to highlight their influences on the methodological approaches of those I interviewed, and to discuss the influence on them of the cultures and structures surrounding the field of arts and dementia. The importance to evaluation of some of the non-epistemic values relevant to arts and dementia was also discussed. In this concluding section, I present examples from my interviewees' experiences of evaluation to demonstrate where adoption of some of the key epistemic values of the prevailing evaluation paradigm had practical repercussions for evaluation practice, and for the roles and relationships involved. The discussion also serves as a useful reminder that evaluation is never just a technical activity, and that methodology does not simply mean the application of method: both involve individuals who possess agency, form relationships, play roles, make subjective judgements, and have emotional responses to events.

While any number of epistemic values might have been discussed (including those elements identified as limitations in the literature review), the ones I have chosen to focus on are: bias, generalisability, replicability, sampling, and ethical process. Attempts to apply these value principles to evaluation in the context of arts and dementia result in challenging experiences for evaluators and for others involved. They also result in perceived distance, alienation, and asymmetric or unequal power relations between different stakeholders. Practices involving these values are also thought to result in burden being placed on people who may already be under significant pressure.

Maximising objectivity and minimising bias are actions widely recognised as desirable for evaluation, and arts stakeholders described attempts to do both of these. Bias is generally thought of as negative, linked to the making of errors and deviations from good judgement, with origins in cognitive or psychological human processes (Wieringa, Engebretsen, *et al.*, 2018). One of the key ways that stakeholders sought to minimise bias was through the employment of external evaluators. However external evaluators can sometimes, as evaluation theorists have noted (Clarke and Dawson, 1999), be more interested in producing a report than seeing its findings implemented; they may be insensitive to, or unaware of, organisational norms and structures, or simply fail to achieve access to the kind of intimate insider knowledge that would enable their findings to reflect the complex reality of the situations and experiences they are evaluating, or to gain the traction needed to make a difference within them.

Interviewees provided examples of arts partners breaking academic ‘rules’ or ‘protocol’ including those governing academic authorship and the way in which claims are made for the results of a study – both of which speak to things that academics generally consider of value, namely that individuals should receive credit for their work, and that their labour should be recognised. As Latour and Woolgar have suggested, ideas of credit and credibility play a central role in scientific life and in the construction of scientific facts (1979).

Both academic evaluators and arts partners interviewed sometimes engaged in evaluation with the ancillary hope of getting academic publications out of the endeavour – another marker of the value of their work, although one that involved them in issues around conflicting interests, ethics, and principles of anonymity. Few arts organisations are likely to possess the capacity to critically appraise an evaluation’s findings or analysis, whether these are quantitative (Oman and Taylor, 2018) or qualitative (Brown, 2010) in nature. Perhaps this is one reason why arts partners sometimes struggled to understand how to translate the findings of an academic evaluation into something that would be useful for them in practice.

While the promotion of empirical studies demonstrating increased methodological rigour (control groups, the use of validated measures, randomisation, etc.), may be one way to answer to these limitations, in practice, evaluators may have little say over any of these aspects. Furthermore, the introduction of technical method fixes whose aim is to improve a study’s results in relation to concepts such as reliability, validity, or generalisability might involve re-structuring programmes in order for their evaluation to align more easily with an ideal (but unreachable) standard. Epistemologically speaking, while increased technical rigour might increase generalisability, it tends to come at the expense of contextual validity (Gartlehner *et al.*, 2006). As Tobin and Begley (2004) have noted, the standards we use to ascribe quality in research have both an ethical and a political aspect. And, as evaluation stakeholders interviewed for this research suggested, not only might it prove prohibitively expensive and resource-intensive to conduct more rigorous evaluation, it was also felt to be ethically problematic by some: a ‘contorted’ intervention might reduce a programme’s impact or reach and also its real or perceived benefits to beneficiaries, funders and commissioners, and to the delivery organisation.

Most evaluations discussed by my interviewees had been commissioned as summative investigations of existing programmes. Evaluators reported having little control over the arts activities they were evaluating, the contexts in which they were delivered, and the participants involved. As in the published literature, an inability to generalise beyond the existing case was cited as a limitation of the results and their usefulness. The small size, variability, and instability of participant groups is a consistent theme in discussion of the challenges of evaluation.

The prevalence of opportunistic samples, a lack of control over who attended sessions, and patchy attendance (because attendance was generally a choice, something that makes ethical, but not epistemic sense) all created difficulties for data collection. Multiple variables were thought to have affected the engagement of participants in an arts activity, including their previous experience, personal characteristics, illness, increasing frailty, cognitive decline and mortality, as well as the natural ups and downs of life. In many projects little was known about participants' health and wellbeing or exposure to the arts prior to the arts activity. Evaluators also described being dependent upon caregivers – both family and professional – for information, and for their support and encouragement of participants, as well as for ensuring that participants attended sessions or contributed evaluation data when needed. This had led to problems where carers acted as 'gatekeepers'. This might mean they decided that an arts activity would or would not suit a particular person, or had let other considerations, such as a desire to make a care home appear busy or colourful, determine an individual's attendance.

The needs and values of the evaluator (to have a stable sample or to know in advance who might be attending, for example) were clearly not always aligned with those of others holding a stake in the evaluation. Additionally, epistemic values learned during professional training or adopted in alignment with hierarchical structures such as those in care settings or hospitals, were shown to have affected the perception and the reporting of the effects of arts practice in those settings. In overcoming such issues for evaluation and research in care settings, researchers have been recommended to be attentive to the ‘philosophical underpinnings’ of individual care homes and to ‘the politics of hierarchy’ within a home and between it and other health professionals, when undertaking research (Goodman *et al.*, 2011).

Clear description of an intervention was described by some of those interviewed as one of the ‘bottom lines’ of good evaluation. This is supported by guidance around quality assessment of scientific studies. It can help evaluators to understand what constructs it may be opportune to measure and how measures and tools should be applied. It can also help to ensure that an intervention and its evaluation will be replicable, hence improving generalisability and supporting future evaluation of similar programmes. Having a stable model that can be applied in different contexts is generally a good thing for evaluation. However, an emphasis on process, improvisation, and flexibility in working creatively with people with dementia was considered a key factor for effective or ‘quality’ practice – and benefitting participants – as has already been discussed above. It has been argued that it is the intrinsic unpredictability of arts and creative initiatives that gives them their capacity for impact on participants, meaning that it might never be possible to apply them ‘to social problems like a tablet’ (Matarasso, 1996, p. 21). The character, skills and experience of artist facilitators provided further unpredictable elements; sometimes activities involved two or more artists working together, increasing this unpredictability. In addition, health and care settings had their own routines and systems to which the arts activity might be peripheral, and evaluators found that these routines impacted upon them, upon the artists and upon participants, and it was difficult to work within and around them.

Problems associated with the application of research ethics processes to arts and dementia evaluation projects were also described. Ethical processes, formal or informal, are important in supporting interaction with individuals lacking the capacity to give informed consent. In the UK, ethics review processes govern the conduct of research, including some evaluation research, since while service evaluation may not require formal ethics review, it may be necessary in certain clinical settings such as NHS hospitals. Researchers also considered ethics review essential where they had ambitions to disseminate the findings of evaluation through academic publication. Evidence that a study has passed through an ethics review process provides another marker of its credibility.

While those I interviewed recognised the need for ethics processes, in practice these could be compromised by a lack of time and resource. Time, familiarity with the person gathering consent, and attention to context have been recommended for proper conduct of consent processes with persons with dementia (Dewing, 2008). Evaluation activity was itself held up when insufficient time had been set aside for ethics review in a project plan which prioritised project delivery. Stakeholders found that complex and formalised consent and information processes, often based on templates used in medical contexts, were off-putting for people with dementia taking part in arts activity and could also erode relationships with care staff or artists. Faced with having to ask for consent at moments when it felt inappropriate or actively harmful, such as at the point at which a person with dementia was being admitted to a hospital, evaluators sometimes rightly decided that the evaluation (and the arts activity) was of less importance than respecting the feelings of a carer or patient in the moment. It was also suggested that a standardised intervention, such as might be required to fulfil the requirements of an ethics protocol, might not fulfil the varied needs of participants, or might entail a compromise in quality, and therefore their experience. And, as one interviewee suggested, assessment of capacity conducted outside the context of an arts activity might be failing to reflect a more nuanced ethical position which recognises the enhanced capabilities of a person with dementia in particular situations (Dewing, 2008).

The examples above describe situations in which epistemic concepts of high value are ‘trumped’, diluted, or otherwise compromised by those that have little to do with criteria of value relating to how knowledge is gained. My interviewees’ experiences of evaluation were full of methodological compromises (and sometimes conflicts) of these kinds. Indices of these can be traced within the relationships and patterns of communication they described.

Intersectoral and sometimes cross-disciplinary relationships and collaborations (between individuals representing hospitals or care homes, arts organisations and different academic or evaluation research disciplines and communities) are characteristic of the arts and dementia field. My findings provided examples where these relationships and collaborations are perceived by those within them as sometimes unequal, affected by assumptions about the purposes of evaluation, by financial concerns and drivers, and also by a lack of clarity as to the needs of the different partners and what they stand to gain from or can contribute to the evaluation process. Difficulties between evaluation and arts partners are also sometimes attributed to misunderstandings around language.

Interviewees reflected on the effects of a lack of time and resource for building partnerships and understandings in the early stages of a project or before it had begun. They commented on cultural misunderstandings about the needs of the care sector, providing examples of projects in which arts partners and evaluators had failed to communicate the benefits of both arts activity and evaluation to care sector partners. There were further examples that suggested communication difficulties between arts partners and the funders of activity. For example, several reported particular difficulties in complying with local authorities' requirements for evaluation. The director of one arts organisation reported consulting with a more experienced provider in order to be able to communicate with their local authority funder with confidence about what it felt were their excessive and inappropriate evaluation demands. Arts partners had also experienced difficulties because of the conflicting requirements and philosophies of multiple funders contributing to a single project.

Characterisations of evaluation as 'burdensome', additional, or marginal to the arts activity itself may reflect a perception that evaluation and the people who conduct it are of lower status than those involved in more core activities (Davies and Heath, 2013), including delivery of arts activity itself. In the music therapy context it has been suggested that 'mismatched' understandings about the purpose of evaluation (outcomes measurement in this case) can affect relationships between therapists and researchers (Spiro, Tsiris and Cripps, 2018). The time and effort involved in being the subject of an evaluation, or being called upon to support it, may also be under-valued or under-estimated by evaluators. As some of those I interviewed suggested, evaluators walk a difficult tightrope between relying upon and exploiting those people who provide them with their data (section 5.4.3).

Where evaluation results seem of little practical use to artists or health and care practitioners, evaluators may be rightly concerned that evaluation activity is viewed as a threat or considered in some way antithetical to the goals or practices of arts practitioners or cultural organisations. Feelings of intimidation, distance, and a sense of lack of ownership over evaluation tools and methods were reported by some arts practitioners. The consequences of this may be to further cement a dichotomy of knowledge that exists between the perceived (observed or evaluated) and actual experience of the arts activity. Experiential knowledge is already at risk because of the problems evaluators have with accessing the subjective worlds of participants with dementia. As a result, realities at the 'sharp end' of experience may not always be well-understood, and changes or improvements made to practice may be mis-targeted, as has been noted to be the case in relation to change in complex healthcare systems (Braithwaite, 2017, 2018).

Interviews for this thesis have shown that evaluative work in arts and dementia entails careful processes of engagement and negotiation. Evaluation theorists Guba and Lincoln (1989) have described a vision of the role of the evaluator as one of 'mediator' or social change catalyst rather than 'describer' or 'consultant'. Some evaluation projects referenced people who, in various ways, moved between evaluation partners because they had knowledge or experience that bridged disciplines and sectors. Often these were artists, facilitating relationships between evaluators and participants with dementia; it should be noted that evaluators could be reluctant to capitalise on the 'intimacy' of these relationships because of concerns about objectivity and potential for bias.

The arts may be a way of 'bringing people into relationships' (Basting, 2014) but what this means for evaluation methodology needs further exploration (Schneider, 2018). It is possible that such concerns might cease to be an issue if different evaluative models were applied, or if the potential for health, care, or dementia-related go-betweens was the subject of greater focus in evaluation planning.

This leads us on finally and perhaps most tellingly, to a relationship that was rarely discussed by my interviewees, except in relation to communication difficulties – that between the evaluator and participants living with dementia. As has already been established, communication with people with dementia can be fragmented, difficult, and unequal – interviews and self-reporting are problematic, and proxies sometimes need to be relied upon. And yet, the agency of people living with dementia, including the choices they make about whether or not to engage with an arts activity, and when and how, as well as their motivations for participation, is a key factor complicating any causal claims that evaluators might endeavour to make as the result of evaluation. The nature of dementia – a condition that can remove or appear to remove agency from those who are living with it – necessarily complicates evaluation. It increases the difficulty for evaluators who wish to account for the values that are important to those who live with the condition in their work. These values, along with the relationship between evaluation participants and evaluator, deserve closer and more detailed attention.

In situations in which there is inequality of access, knowledge, or cognitive capacity, a key ethical challenge is to find methods through which an evaluator can enable rather than disable participants' unbiased engagement with an evaluation. In evaluation involving the kinds of complex social systems in which artists, arts organisations, health and care practitioners, and people affected by dementia find themselves playing roles and developing relationships, it may be necessary to develop methods that situate 'communicative competence' at least equally with 'technical competence' (Sanderson, 2000).

Although the challenges of cross-, trans-, or multi-disciplinary working are many, it has been acknowledged that arts and health evaluation and research will benefit from embracing these approaches (Newman *et al.*, 2016). However, this exploration of the experiences of evaluation stakeholders suggests that it may not be realistic to propose that values are easily shared or that tensions are not going to have methodological effects. The discussion above about value compromises suggests that it might not be sufficient to trace methodological difficulties to failures in dialogue, translation, or communication across and between sectors and disciplines. As a result, it is necessary – and not simply helpful – to identify the different kinds of value that different partners associate with evaluation and with arts activity, and to develop processes for evaluation through which these values can be more equitably and transparently embraced, acknowledged, and balanced against each other, even when they cannot be shared.

Discussion in the previous chapter suggested that evaluators face difficulties because of how the act of ascribing meaning to the engagement of people with dementia in arts activities can entail the application of both epistemic and non-epistemic values. This may be particularly problematic when evaluators aim for methodological rigour, finding that this steers them into uncomfortable compromises and difficult relationships. The next and final chapter reflects, summarises, and signposts some solutions for future practice, as well as discussing the strengths and limitations of this research.

7 UNFASTENINGS

> O time, thou must untangle this, not I.
> It is too hard a knot for me t'untie.
>
> William Shakespeare, *Twelfth Night*

The aims of this research included a desire to signpost solutions. The origins of the word *solution* lie in the Latin verb 'solvere'. This can be translated as 'to unfasten' or 'to free'. What I offer in this chapter are solutions only in the sense that they might serve, alongside other research, to unfasten slightly the knots that we tie ourselves up in trying to solve the methodological challenges of this work. These *unfastenings* are informed by the contents of the previous chapters; they are thus grounded both in the challenges reported in the literature and the experiences of evaluation stakeholders. The term I use also provides a metaphorical link to those tensioned warp and weft threads within the fabric of methodological challenge described earlier (section 2.6), to the theoretical influence on this study of the idea of method 'assemblages' (section 3.3), and to the tangled connections between research and evaluation, policy, and practice discussed throughout the preceding chapter.

7.1 Overview

As I have found through my interviews for this research, people involved in the arts and dementia field are generally quite open about their belief in the value of the work they do and its capacity to improve the lives of people living with dementia. There is less articulacy about quite how this improvement is achieved and what it looks like in practice (section 5.3.2).

The prevailing discourse around evaluation in the field (even if not explicitly stated) is one that presents its duties as two-fold: to demonstrate that this improvement is real, and to do this in a way that is going to provide support for its continued facilitation. Application of understandings from the field of cultural policy rehearsed in previous chapters indicate that such assumptions, alongside the continued debate around the value of arts and culture, may be leading to some of the considerable methodological difficulties evaluators face.

However, perhaps the most fundamental source of methodological difficulty for the field is the chronic and degenerative nature of the condition of dementia itself. Over time it sadly removes agency and disrupts the cognitive and communication processes of those who are living with it. Although it seems highly possible that arts activities may hold or help to hold some symptoms at bay (reducing the rate of decline) or alleviate or distract from them in the moment (enriching or improving quality of life or contributing to some other form of wellbeing), few, if any, would claim that arts activities will cure or reverse the course of the disease. Evaluation designs and the measures used should, therefore, be adequate to the job of assessing the value of an arts activity, given this trajectory. So, for example, measuring a reduction in the rate of decline is one thing, but an evaluation must also be capable of assessing the value of this reduction to the individuals concerned. If we suggest that an arts activity enriches a person's quality of life, a baseline or set of criteria has to be identified against which the value added can be measured.

Further, for arts and dementia evaluation to play the role its advocates would like, the methods it uses to make a judgement on the value of arts engagement need to be adequate in the two dimensions of epistemology and values.

This means that these methods do need, in line with the dominant scientific paradigm, to enable things like measurement, confirmation of hypotheses, theorisation, and so on as successfully and as credibly as possible. The good news is that despite multiple practical challenges, it is increasingly evident that this may be, methodologically possible. However, to benefit and – further – to empower people with dementia and to justify a place for the arts within therapeutic or clinical applications and interventions related to dementia care, I suggest that the concepts and claims associated with the benefits of the arts also need to be value adequate. With Alexandrova (2012) I use the term ‘value adequacy’ to mean the ‘fittingness of the nonepistemic values presupposed in [a given scientific] claim to the values at stake in the context in which this claim is made’ (p679).

Put simply, this means that when we research and evaluate arts and dementia activities, the benefits that we believe might accrue from the arts should be a reflection of the things that those most closely concerned would agree to be valuable. In other words, it is important that they incorporate a widened focus to include things like individual identity and agency, relationships, community, and rights. If the two dimensions (epistemic and value adequacy) are not in balance, the result for evaluation is likely to be tensions and unsatisfactory compromises (section 6.4). To resolve these kinds of tensions, I suggest that evaluators and evaluation stakeholders will need to justify, describe, and explain their methods in ways that allow an equitable relation between the two.

In practice, this will mean evaluators working in collaborative and participatory ways and using methodological practices that acknowledge the multi-disciplinary nature of the field. Study objectives, designs, measurement tools, and methods need to be sensitive to the symptoms of dementia and respond to those needs and values that are specific to the condition of dementia and to the individuals involved, as well as those that relate to arts engagement (section 2.6 and section 6.3.1). Continuous reflective attention to how multi-disciplinary stakeholders locate and communicate meaning within the processes of arts engagement is also necessary; this has been significantly aided by recent research suggesting a common language of classification and description for the component parts of an arts and dementia ‘intervention’ (Cousins, 2018).

7.2 Signposts towards a values-informed approach for evaluation practice

In this concluding chapter I provide some signposts for future evaluation practice. The starting point for this is the suggestion that evaluators should adopt a values-informed approach to evaluation of arts and dementia activity. This aligns with previous calls for person-centred research practice as a result of observations that a focus on person-centredness in nursing care practice has not always translated into research methodologies. This focus highlights a perceived divide between researchers (those who strive to know) and practitioners (those who act) (McCormack, 2003). I provide a sketch outline of some of the key (learnable) skills that might be applied to enable this.

Pointers follow which would align with and support a values-informed approach. These are: the application of a complexity lens to the problem; the need for collaborative and multidisciplinary working, and; methods innovation that gives evaluators better access to, and allows all stakeholders to benefit from, the experiences of people living with dementia and of artists working in the field.

Finally, to be useful, solutions need to reflect the spectrum of kinds of evaluation practised in the field. Technical fixes involving complex, rigorous, or expensive application of method are not going to cut it for more than a small number of evaluators. Therefore, I also make a plea for recognition of the limitations for all, but particularly non-expert evaluators in the field resulting from a lack of capacity and resources. I propose a critical attitude towards ‘use’ (evaluation utilisation) in the light of a values-based approach; this might involve reframing arts and dementia activity itself as an ethical practice.

These pointers are offered here in a spirit of exploration, as directions that need to be critically engaged with, refined, and adapted in practice.

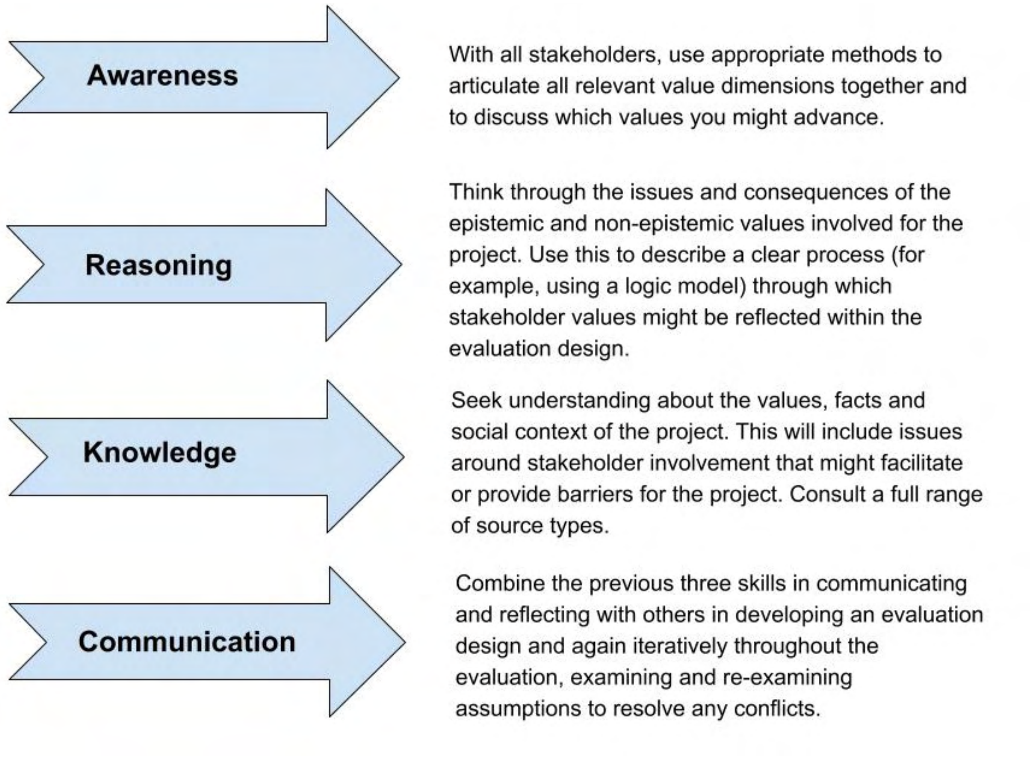
7.2.1 A values-informed approach to evaluation

In clinical care, values-based practice offers a different but complementary response to evidence-based practice (Fulford, 2011). In recognition of the complex and often conflicting nature of values, and the fact that the values of others are not transparent to us, and our own are often unacknowledged, its proponents have described a set of four learnable clinical skills that can be used to inform the making of balanced evaluative judgements in individual clinical cases. These have been combined with six other key elements to describe a coherent approach through which it is suggested that value differences can be made a positive resource for practice (Fulford, 2011, pp. 979–983).

Fulford's approach has recently been adapted for application in a manifesto to support decision-making processes in dementia care (Hughes and Williamson, 2019). My proposal is that they could be further applied in forming a set of principles to underpin a values-informed approach to evaluation of arts-based activities for people with dementia. I have used the phrase 'values-informed' rather than 'values-based' in order to emphasise the critical nature of the approach. However, as with values-based practice, such an evaluation approach could include attention to trainable skills, guidance to support professional relationships, and the development of appropriate evaluation environments. McCormack's principles and outlined 'necessary and sufficient conditions' for person-centred research practice, along with the factors that are required in order to enable those conditions, might provide an important further influence in shaping the approach (2003, pp. 185–187). A values-informed approach, although focused around issues of inclusion of people with dementia, would not stop there. Values-based practice was originally developed to support the translation of evidence into good care and balanced decision-making between clinicians and patients. Therefore, the application of its principles would strengthen links with evidence-based practice – and could be applied irrespective of the methodological framework chosen for evaluation.

Figure 4 below provides a sketch outline of how Fulford's first four trainable clinical skills might be applicable within an evaluation context.

Figure 5: Four practice skills to inform a values-informed approach in evaluation



Awareness	With all stakeholders, use appropriate methods to articulate all relevant value dimensions together and to discuss which values you might advance.
Reasoning	Think through the issues and consequences of the epistemic and non-epistemic values involved for the project. Use this to describe a clear process (for example, using a logic model) through which stakeholder values might be reflected within the evaluation design.
Knowledge	Seek understanding about the values, facts and social context of the project. This will include issues around stakeholder involvement that might facilitate or provide barriers for the project. Consult a full range of source types.
Communication	Combine the previous three skills in communicating and reflecting with others in developing an evaluation design and again iteratively throughout the evaluation, examining and re-examining assumptions to resolve any conflicts.

Outlining how these skills might be developed within an evaluation is beyond the scope of the current project. However, the following sections flesh out some of the ways in which evaluation might address challenges around professional relationships, creation of the environment that might enable practice, and they place collaborative and multi-disciplinary working alongside person-centred practice at the heart of evaluation.

7.2.2 Applying a complexity lens to the problem

Applying a complexity lens to the problem might be one way to bring the diversity of values involved to the fore. The arts activities described by researchers and evaluators take place in hospitals, care homes, community settings, cultural venues, and private homes. The settings are described by evaluators and researchers as being complex and open. Evaluation is generally conducted of existing programmes. In natural contexts such as these, evaluators have little control over the conditions, content or setting of activities being evaluated (section 2.5.3, section 5.4.1, and section 6.4). Organic, spontaneous, and creative in the moment variability is often described as integral to artistic practice in the field (section 5.4.6). However, resulting problems with operational definitions are sometimes thought to hamper attempts to make interventions replicable and the changes associated with them measurable. As I have shown, for stakeholders in evaluation, there are concerns that the application of methodological parameters or principles in evaluation will ‘contort’ an activity being delivered, reducing the usability of evaluation findings and the activity’s value in practice (section 5.4.1, section 6.4).

Given such characterisation of the settings and the activities that are delivered in them, one potential solution may be to explore evaluation approaches that draw on ideas of complexity, and in particular, complex adaptive systems (Hawe, Shiell and Riley, 2009). Such systems have been defined as ‘a collection of individual agents with freedom to act in ways that are not always totally predictable, and whose actions are interconnected so that one agent's actions changes the context for other agents’ (Plsek and Greenhalgh, 2001). They have fuzzy boundaries and feature other systems and collections of agents nested within them, whose actions and interactions can lead to emergent and unexpected phenomena.

Conducting research or evaluation from this perspective might be termed applying ‘a complexity lens’. Such a lens would lead us to vary our focus. We might, for example, want to consider a person with dementia taking part in a storytelling and music activity as: one resident living among others in a care home that is itself one in a group of care settings, and managed by a charity with a particular culture and philosophy of care; a patient with symptoms of a particular form of dementia who receives pharmacological and non-pharmacological treatment as part of her care by healthcare practitioners; a much loved member of an extended family with a distinctive set of relationships and responsibilities; an immigrant living in the UK who is a member of a religious minority ethnic community with specific rituals, traditions, and history. In addition, as we know, the artists facilitating arts activity will bring their own histories and experiences to the event, as will the care staff supporting the participant. As Audre Lorde put it, none of us live single issue lives (1986, p. 133 (2019)) and we shouldn’t be treating people with dementia as if they do.

However, effects felt in all these interconnecting systems and for any and all the individuals involved are difficult to resolve or capture using conventional evaluation methods. It is not possible to extend our analysis to all of them, and the point at which we cut off our focus is always going to render some things invisible. To facilitate and understand change within complex systems therefore, it may not be enough to increase the knowledge we hold about agents within them, to improve models or theories, or the quality of our methods of data collection, analysis and reporting in positivist terms. Plsek and Greenhalgh suggest these are no substitutes for a shift of focus to the bigger picture in order to allow careful observation of the patterns of behaviour that occur within those systems (2001, p. 627).

As an example of the kind of interactions such a lens might reveal, in one hospital setting, the director of an arts organisation I interviewed reported that an arts project had been able to bring together different clinical specialists including speech and occupational therapists, a nurse, and a nursing manager, all discussing the needs of a particular patient; but she noted, this conversation might not have happened without the arts project: ‘it was shocking to me that those four people didn’t all sit down and talk about each of the residents on a regular basis’ (Sam).

To understand the extent of, and the barriers or facilitators to any changes the arts activity ‘event’ has caused (or not), an evaluator might – for example – aim to track the traces and markers an activity has left at key points within the different interacting systems and for their agents. She might seek to observe alterations to the structures of relationships and the creation of new roles and resources, note examples that might suggest boundary-crossing behaviours (Daykin, 2020), or indicate that an arts activity is enabling, displacing or otherwise altering other activities – to either positive or negative effect. Careful reflection on where we place our boundaries for observation will also be crucial (Midgley, 2003).

Such approaches would foreground the kind of contextual factors that provide the normal conditions in which arts activity is practised and which – in a trials-based model for evaluation – are sometimes viewed as contextual confounders to be eliminated. They might move evaluation away from short summative impact studies towards evaluations over longer time frames. A focus on implementation would also help to answer questions about scaling up, replication, and sustainability that are likely to be of interest to policymakers and those who might commission arts activity.

In healthcare contexts this could be achieved through the adapted application within studies of conceptual frameworks and theoretical approaches such as the consolidated framework for implementation research (Damschroder *et al.*, 2009) or Normalisation Process Theory (May *et al.*, 2011), depending on the particular question under consideration and whether its focus is on the contributions of the system or the actions of agents within it. In such explorations it will be useful to understand context not as the place or physical environment in which an activity takes place, but as a process whose forces can be harnessed for change (Axelrod and Cohen, 1999; May, Johnson and Finch, 2016).

Mixed methods are recommended for evaluation of complex interventions and would be a facet of such work. Specific applications might include process evaluation studies, including those that are qualitative (e.g. Bridges *et al.*, 2017, an example from outside the arts and dementia field) and which are aimed at capturing the dynamics and fluidity of change processes. As a first step to an implementation study, ethnographic and ecological approaches would be useful in understanding an arts activity taking place in both institutional and less formalised social systems, such as private homes or the community.

Theory-based evaluation offers an alternative to experimental or impact studies and promises a generative rather than simple cause and effect understanding of causality. This is aligned with an appreciation of complexity (Galloway, 2009; Pawson, 2013) and an understanding of the ways in which a program of activity are shaped by the actions of individuals and by context. Realist approaches are increasingly applied to the evaluation of complex healthcare interventions because they promise to provide an understanding of just these factors – exploring what works, for whom, and in what circumstances (Pawson, 2013). They have recently been advocated for studies of arts and health as an antidote to the ‘outcome-driven, quasi-experimental approaches imported from clinical trials’ (All Party Parliamentary Group on Arts Health and Wellbeing, 2017, p. 40). A combination of critical realist or realist evaluation and arts-based methodologies also has potential to provide a model for knowledge transfer (Kontos and Poland, 2009).

It will be important to seek more pragmatic ways of applying the methods and understandings that accompany such work since use of any and all these approaches will be limited by the resources available to fund evaluation, as well as by the skills and knowledge of evaluators. In order to apply evaluation approaches deriving from realist methods or implementation science, most arts organisations would need to take on the role of service delivery partners, working in partnership with academic or other evaluation consultants as part of major funded implementation studies.

7.2.3 Multidisciplinary and collaborative approaches

Since the findings of this study would suggest that it is neither practical nor desirable to eliminate value from evaluation of arts activities for people with dementia, it is evident that, as an activity, the work requires the input of a diverse community of interests. It is unrealistic to expect any evaluation process to reflect a single shared epistemology or set of values and not to experience methodological challenges that may compromise its results. The solution is to acknowledge and seek ways to make positive use of the value diversity we find within teams, and across disciplines and sectors.

It would be naive to suggest that this might be easy. However, my interviewees for this research described experiences of positive and profitable collaborative relationships between sectors, disciplines, and stakeholder groups working on evaluation. These collaborations allowed evaluators to access important practical support in areas such as the recruitment and retention of evaluation and project or activity participants. Collaboration was considered successful where roles, responsibilities, and aims were clearly understood, complementary skills and knowledge shared, regular meetings and governance processes established, but where evaluation partners retained some independence and a sense of the value of their own contributions to the project and what they wished it to achieve. As several of my interviewees commented, evaluation activity is sometimes negatively positioned and accepted as a ‘trade-off’ in a deal through which – for example – a care partner receives arts activities and can present a positive image of their care setting to potential ‘customers’, and an arts organisation receives material for advocacy:

> […] what I’ve seen is some quite good teams where you’ve got successful arts projects that are in quite big organisations that have teamed up with researchers and done some good basic science or research and then used that as part of an advocacy kind of programme. […] So they aren’t trying – the artists aren’t trying to be the scientist and vice versa.
>
> (Naomi, academic researcher)

In developing promising approaches for future practice therefore, evaluators may wish to borrow from progress in other fields in which it has proven necessary to work across disciplines and sectors and to engage community members to solve complex or ‘wicked’ problems related to the environment, international and community development, and the wider field of public health. Such ways of working go beyond mere ‘collaboration’. Here pluralistic frameworks have been recommended for research and evaluation activity with multiple stakeholders that is not tokenistic, and that can accommodate multiple epistemologies without creating a poorly crafted and ‘stapled together’ product that, in the end, is likely to please no one (Miller *et al.*, 2008).

For arts and dementia evaluation purposes, this could entail approaching evaluation projects in quite a different way and for perhaps subtly different ends. For example, by taking an integrative applied approach (Bammer, 2013) attentive to questions of value, evaluation might explore the question of how feasible it is that an arts activity will benefit people with dementia rather than starting from the assumption that it does. It might start with scoping the problems that the arts are seeking to solve from the perspectives of agents within the different systems involved. It would identify the disciplinary and sectoral expertise and stakeholders that might contribute to a solution and – together – set boundaries to, frame and agree the evaluation problem, using the discourses of the systems concerned. It might then consider the place of values within the evaluation and the activity itself, and seek collaborative means to harness and manage the differences in perspectives. Such an approach, positively informed by an appreciation of complexity (section 7.2.2.), would also have the benefit of concerning itself with and contributing to the planning, development, and implementation of programmes of arts activities.

Trans- cross- or inter-disciplinary and multi-sectoral working is notoriously difficult, with multiple institutional and other barriers to its practice (Lélé and Norgaard, 2005). To overcome these, one recommendation is that attention be paid to the disciplinary and institutional characteristics of the evaluation teams working with arts organisations. Some university departments may – on their own – be poorly equipped to enable the kind of innovative strategies needed to solve arts and dementia evaluation problems.

Latour and Woolgar (1979) showed that we might want to view the scientific system as a social structure within which individual researchers are enabled or constrained, with particular kinds of activity and relationships stimulated or inhibited by the system in which actors will be motivated by – for example – a desire to build reputation and peer recognition. Commissioned academic consultancy in which expertise and a knowledge product (a report) are exchanged for (sometimes admittedly negligible) financial gain, publication, or the promise of increased *impact* does not necessarily provide a good model for collaborative values-based evaluation practice.

One recent research project, involving a group of researchers, artists, and media professionals from a range of backgrounds attempted to shake up traditional disciplinary location and identities by moving itself, for the duration of its two-year exploration of creativity and the dementias, out of an academic institution and into the 'Wellcome Hub' – a space specifically designed to encourage interdisciplinary research practice (Camic *et al.*, 2018). Following on from this endeavour, it may be profitable to explore approaches that 'decentre academia' and help stakeholders to '*un*learn' practices (Alonso-Yanez *et al.*, 2019), in order to emphasise co-creation or co-production of knowledge instead of its simple transfer or translation from academic to non-academic stakeholders.

Nowotny and colleagues introduced the concept of knowledge production taking place in the *agora* – describing a public space in which knowledge is produced and traded and its various producers can jostle productively with their 'publics' (Nowotny, Scott and Gibbons, 2003). A proposed strategic centre for arts, health, and social care (All Party Parliamentary Group on Arts Health and Wellbeing, 2017) could provide a (virtual) space for this kind of work – although some might remain wary about its advocacy intentions, and dementia would be only one of the areas with which it is concerned. Research intermediaries modelled on those like the Alliance for Useful Evidence could provide links between the knowledge produced and policy (Breckon and Dodson, 2016). Boundary organisations, including the research councils, ethics review boards, and the national Arts Councils, have an important but difficult role to play in providing resources and guidance that can balance the interests of science, the public (who in this case include the service delivery organisations as well as people affected by dementia), and politics (Guston, 2001). Their representatives also might benefit from taking part in collaborative 'social *un*learning' processes with project stakeholders designed to shake up their thinking around procedures, systems, and processes that they might previously have considered essential. While we might want to be wary of the idea that formal reflective thinking methods are adequate on their own to critical tasks involving relationships, boundaried knowledge, and multiple perspectives, facilitated dialogue and discussion around philosophy and values between stakeholders (Eigenbrode *et al.*, 2007; O'Rourke and Crowley, 2013) could usefully contribute to application and training around the four skills outlined above (section 7.2.1).

7.2.4 Methods innovation

Along with the inclusion of a systems perspective and attention to governance and process, 'a creative approach to research focused on improving human experience' has been recommended as a key principle that will contribute to successful co-creation or co-production of knowledge (Greenhalgh *et al.*, 2016). If we agree that the field of arts and dementia should embrace and acknowledge the role of prudential or non-epistemic values in the production of knowledge about its practices, creative approaches could be considered key to the process. Attention to methods that can better access the views and experience of people with dementia, an understanding of the contribution of narrative, and attention to the mechanisms underpinning artists' reflection *in* or *on* action (Schön, 1983) are central to the task of repositioning arts activity with people with dementia as an ethical and ‘careful’ practice (Tronto, 1998). This is one in which the needs, feelings, and perspectives of both participant and practitioner are recognised and their values balanced and accounted for.

There is less research in the field of arts and dementia that attempts to access the subjective experience of people with dementia than there is focusing on the needs and experiences of caregivers or outcomes that relate directly to care or care contexts. We might trace the reasons for this to the effects of excess disability (Sabat, 1994), malignant social psychology (Kitwood, 1997), or stigma that can prevent people with dementia becoming involved in activity of any kind (Batsch and Mittelman, 2012; Swaffer, 2014). While evaluators and researchers certainly report significant difficulties with all the direct methods conventionally used to access the experience of subjects (section 2.5.3, section 5.3.1), with care, meaningful responses can be elicited using the common data collection methods, maybe even from persons whose cognitive impairment is significant (Sixsmith and Gibson, 2007; Zeilig, 2016). If we wish to do this it will be necessary to fit our methods to people with dementia, rather than ‘misfitting’ them through the process of evaluation (Webb *et al.*, 2020).

One implication of the difficulties researchers and evaluators face is that evaluative activity seeking to be attentive to the personhood of the person with dementia will need to base its enquiry on methods that embrace and empower communication and capability rather than deficit models of cognitive failure or impairment. The arts, and arts-based methods of research, offer refreshing opportunities to do just this. For example, it has been suggested that they may: provide alternative channels for communication and self-expression including gesture and sound (Hannemann, 2006; Gregory *et al.*, 2012; Hughes, 2014; Zeilig, Killick and Fox, 2014) and support participants in ‘saying the unsayable’ (Bartlett, 2015); offer stimulus for dormant abilities (Swinnen, 2016); elicit responses during interviews (Hara, 2011); or, through film, provide a medium that enables collaborative representation (Ludwin and Capstick, 2015) and an opportunity to ‘embrace *diverse points of* viewing’ (Goldman, 2006).

Despite this potential, it has been noted that research and evaluation practice still lags behind dementia care itself in using innovative methods to include and empower people with dementia, and that there may therefore be a need for training and support of researchers to better equip them in doing this (Phillipson and Hammond, 2018). There are increasing calls to explore research methods that can better enable, support, and reflect the differing needs and desires of the diverse community of people affected by dementia and also support their involvement in research (Abendstern *et al.*, 2019; Open Doors and Swarbrick, 2020) and in society (Shakespeare, Zeilig and Mittler, 2019)

Narrative, vignette, and the telling and re-telling of stories about the experience of engaging in the arts are methods that can be used to convey and structure data meaningfully. Some of those I interviewed felt story to be a particularly apt method to use in interpreting the experience of people with dementia, perhaps echoing Hughes' suggestion that we 'approach people with dementia as if they were works of art' (2014, p. 1407):

> So, people with dementia who talk in kind of metaphors or other ways and often really put their finger on what they're feeling in that way that makes sense in the moment…. That has to be documented as a story with that sentence in it, rather than after a session with *'can you give me some feedback about how it was today?'*
>
> (Sam, director of arts organisation)

As has already been discussed in the findings of this study, people often use stories to make sense of experience that otherwise might seem too difficult or complex to interpret.

Releasing reflective practice from the business of providing evidence of impact is essential for development of a more active and critical evaluation practice. The potential for qualitative evaluation approaches incorporating artists' reflections has already been discussed (section 6.3.2); in these reflective learning could be viewed as an active process. This might bring it into alignment with participatory action research methods, for example (Baum, MacDougall and Smith, 2006).

This recommendation is based firmly in the experience of artists I interviewed for who described their own careful reflective practices. For example, one had a well-developed framework of meaningful measurement she applied in sessions, based on experience and an understanding and adaptation of the principles used in DCM, and which she described as a way to share knowledge with others and a means by which to focus observation and discussion. One practitioner interviewed described how reflection was an integral part of delivering a programme in which small groups of professional artists worked with care home residents living with dementia. Immediately before an activity, the artists had a short period of rehearsing, planning, and reflecting together. Afterwards, they would reflect again with care staff or carers, a dementia awareness facilitator, and a project manager who had been documenting the session. All contributed to discussion of how artists and participants had engaged with the activity on the day, including how they felt, as well as what they thought or had observed. The intention was to build a composite phenomenological picture of the session for all those involved, and to use these understandings to inform future practice. Both evaluators and artists can learn and benefit from processes such as these. The understandings gained can contribute to quality improvement cycles. Their incorporation into evaluation would also be helpful in changing damaging perceptions of evaluation among artists as *other* or burdensome.

People with dementia and artists will have stories to tell about their joint engagement in the arts and what it means for them. Narrative research offers a means to capture values and structures of meaning that are shared by groups of people. It has been claimed that – as a method – it holds great promise for reporting and illuminating cultural contexts (Greenhalgh, 2016). Many in the health world are already familiar with the ‘patient story’ as a tool that can build a bridge between clinical knowledge and treatment and individual experience of ill-health. As sense-making devices, stories are open to all kinds of human bias but, accompanied by measures and training to strengthen their credibility with a sceptical audience, narrative methods could make a useful contribution within a values-based approach to evaluation of arts and dementia activity.

And finally, placing engagement in the arts alongside meaningful engagement in other creative and leisure activities might lead us to reframe practice of the arts with and for people with dementia as an *ethical practice* (by which I mean one that deals with what it is best to do for an individual in a particular set of circumstances) rather than a science, for the purposes of its evaluation. The same reframing has been suggested for clinical medicine, along with the observation that this might make it a requirement for it to be more open to the inclusion of narrative, reflection, and individual experience in our endeavours to improve and evidence it (Greenhalgh, in Jones et al. 2015).

7.2.5 Acknowledge problems of evaluation capacity and resource

Acknowledgement of a problem provides a starting point for change: the arts sector undeniably suffers from a lack of evaluation capacity, resources, and skills. This not only creates barriers to evaluation practice but also contributes to a proliferation of low-quality impact evaluation that, if it does not actually damage the credibility of the work, certainly does little to increase our knowledge about it. Attempts to increase capacity through mentoring and consultancy have been successful within individual projects (Daykin, Attwood and Willis, 2013) although the benefits are likely to be difficult to sustain (Gibson and Robichaud, 2017) and there appears to be little evidence around how successful they might be at an organisational level.

As well as creating an environment in which arts partners can work more collaboratively and on a more equal footing with other stakeholders to deliver values-based evaluation, there remains an imperative to provide arts organisations with the resources and skills necessary to deliver scalable and useful evaluations. These should be dementia-specific, ethical, light touch, and – as an alternative to focusing on impact – might aim to support the sector in improving the quality of the work it does and in communicating this improved quality with confidence to anyone who needs to understand it.

The findings of my study would suggest that artists and arts organisations might benefit from training to help them understand how to involve people with dementia in evaluation, and to align their own needs and cultures with the needs and culture of health and dementia care settings (section 5.2.3, section 6.2.1). Health and care partners and practitioners will also benefit from training that supports them to invest in and become co-contributors to evaluation and research conducted in their settings. Learning and collaboration might be cemented through networking and knowledge exchange opportunities, locally and nationally, that are inclusive of all those represented within the arts and dementia community of interests.

Organisations operating under a consultancy model, such as the Charities Evaluation Service in the UK, exist to support particular sectors to evaluate their work and even offer evaluation training specifically for funders, commissioners, and investors. There may be scope to develop a similar boundary-spanning organisation for arts and dementia, perhaps sited under the auspices of a national or international dementia research initiative. In many ways however, as was hinted earlier in discussing the need for integrated collaborative approaches, and in relation to guidance for practice (section 6.2.3), there is a risk that this would simply be responding to and upholding the status quo, rather than driving change.

Ultimately, systemic change is needed if arts organisations are to be freed from the unrealistic expectation that they need to provide robust evidence of impact resulting from the short, varied, and contextually different programmes of activity they habitually deliver. There are no easy solutions that can enable this. However, a halfway house, suggested by several of those I interviewed, might involve encouraging funders to work towards the co-ordinated collection of data using shared measures and agreed sets of contextual information which, if made publicly available, would therefore support different kinds of analysis across programmes of funded activity over time.

7.3 What use is evaluation?

Use or *utilisation* is a core but hotly debated construct in evaluation (Weiss, 1979, 1998). In the field of arts and dementia it is not one that gets a great deal of methodological attention, beyond discussion of the problems posed by an advocacy agenda. Evaluation is not a neutral activity, it has effects, and if we are going to propose a values-based approach to evaluation and a re-orientation of perspective through which arts activity might be viewed as an ethical practice, then the question of how evaluation – its process and/or its findings – will be used becomes increasingly pertinent.

Research questions are a good guide to research methods. To give an example of how a re-consideration of *use* from a values perspective might have methodological implications, consider that an evaluation client might ask an evaluator to demonstrate the positive change an arts activity brings about for individual participants' wellbeing. Applying a values-informed approach to the evaluation might reveal that the true goal or non-verbalised subtext for implementation of the programme is whether it has the capacity to contribute towards the solution of a more generalised problem in a particular context, for example – that staff in a care home have little time to devote to getting to know the residents who live there and that this is getting in the way of their delivering person-centred care. In such a case, the findings of an impact or outcome evaluation – even if they showed the results the evaluation client was hoping for – might be of limited use, in either epistemic or non-epistemic (or social) terms to its stakeholders. However, an evaluation that identified, articulated, and then proceeded to answer questions of more direct relevance to those stakeholders and included them whilst it did so, might have an effect and be of use during the evaluation process as well as potentially enlightening or playing an influencing role afterwards.

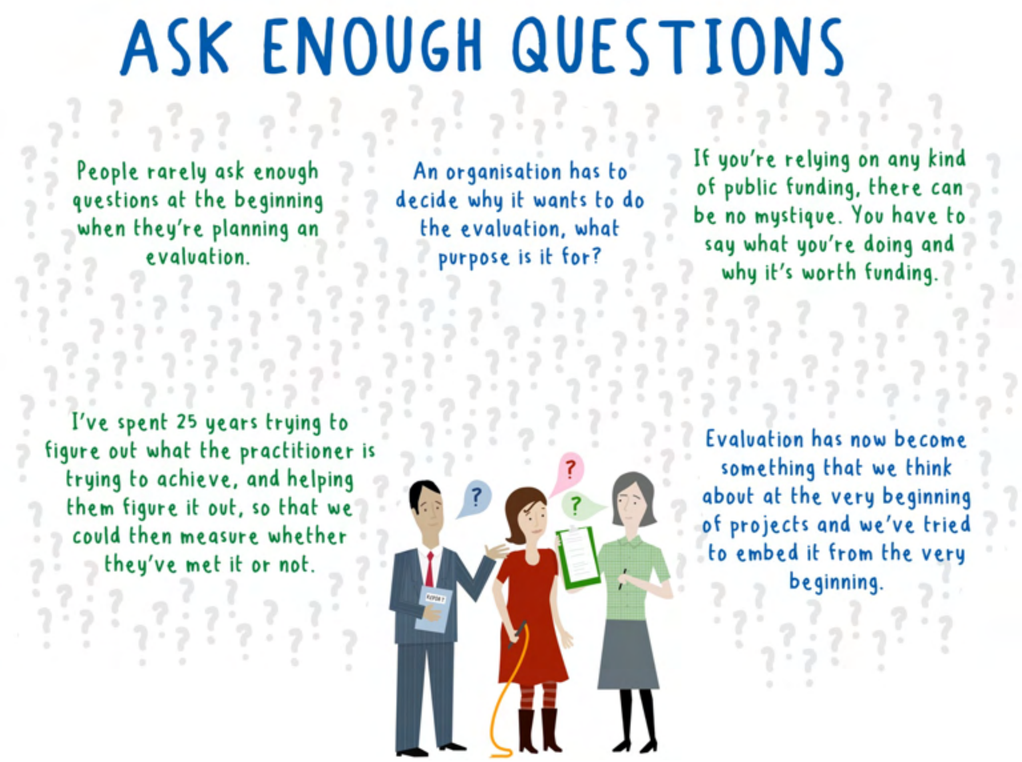
7.3.1 You can't just evaluate for the sake of it

The recommendations for evaluation practice described above were also reflected in the graphic narrative developed to help communicate the findings of the thesis (see section 4.4.7 for the process involved, and section 5.4 for the first part of the narrative). As well as outlining the challenges interviewees faced, this narrative communicates a set of principles, derived from their words, that demonstrates the ways they had found of overcoming the challenges of evaluation and of conducting evaluation that was more directly useful to those involved. As one arts manager put it: 'You can't just evaluate for the sake of it.' These principles are:

- Ask questions.
- Find out the value.
- Get the partnership right.
- Develop a community of practice.
- Be reflective.
- Don't just imitate.
- Make meaning.
- See the success of the project in the context in which it's delivered.

These are illustrated in the images below.

Figure 6: Graphic narrative (Solutions)



ASK ENOUGH QUESTIONS

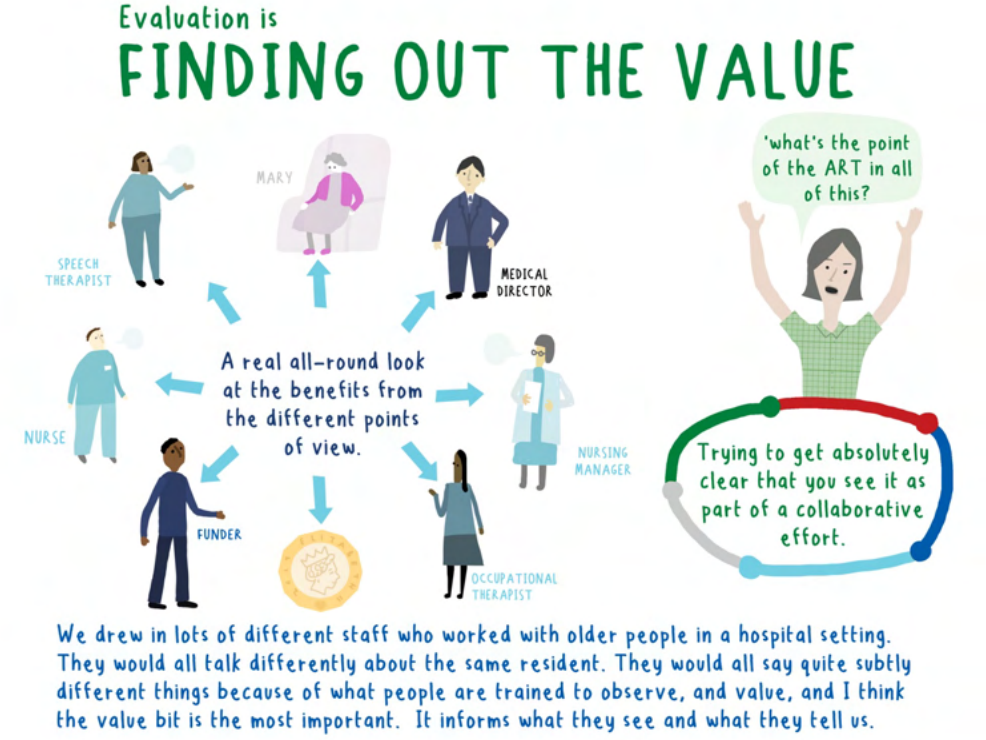
People rarely ask enough questions at the beginning when they're planning an evaluation.

An organisation has to decide why it wants to do the evaluation, what purpose is it for?

If you're relying on any kind of public funding, there can be no mystique. You have to say what you're doing and why it's worth funding.

I've spent 25 years trying to figure out what the practitioner is trying to achieve, and helping them figure it out, so that we could then measure whether they've met it or not.

Evaluation has now become something that we think about at the very beginning of projects and we've tried to embed it from the very beginning.



Evaluation is FINDING OUT THE VALUE

A real all-round look at the benefits from the different points of view.

Trying to get absolutely clear that you see it as part of a collaborative effort.

We drew in lots of different staff who worked with older people in a hospital setting. They would all talk differently about the same resident. They would all say quite subtly different things because of what people are trained to observe, and value, and I think the value bit is the most important. It informs what they see and what they tell us.

GET THE PARTNERSHIP RIGHT

It's multi-disciplinary, you've got to have the programme design and, you've got to have the artist, you've got to have the care staff, you've got to have the researchers, all working together...

...with a clear understanding of each other's roles. So the artists aren't trying to be the scientist and vice versa.

Different disciplines learning from each other.

CONNECT WITH A COMMUNITY OF PRACTICE

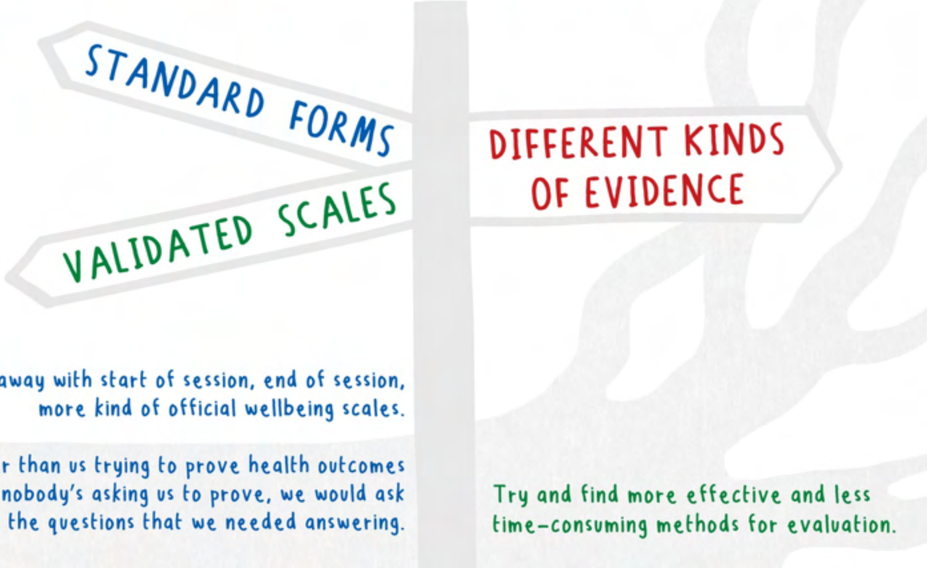
The very best practice in care for people with dementia is the least well prescribed or described or defined because by definition it's more organic,
spontaneous and creative,
in the moment.

I meet a lot of people who are doing things that are very relevant but they're not connecting it with a community of practice. So there's an issue about a silo and the lack of intellectual rigour that that creates.

It's about the relationship that people with dementia have with each other, and also with the care staff, and also with us.

DON'T JUST IMITATE

We shouldn't just accept the standard forms. We've got to better understand the process. We haven't got to just imitate what researchers are doing and automatically reach for validated scales.



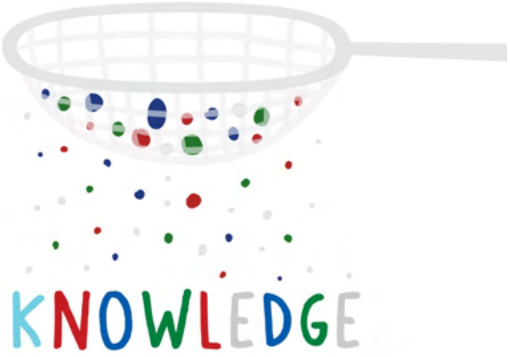
We did away with start of session, end of session, more kind of official wellbeing scales.

Rather than us trying to prove health outcomes that nobody's asking us to prove, we would ask the questions that we needed answering.

Try and find more effective and less time-consuming methods for evaluation.

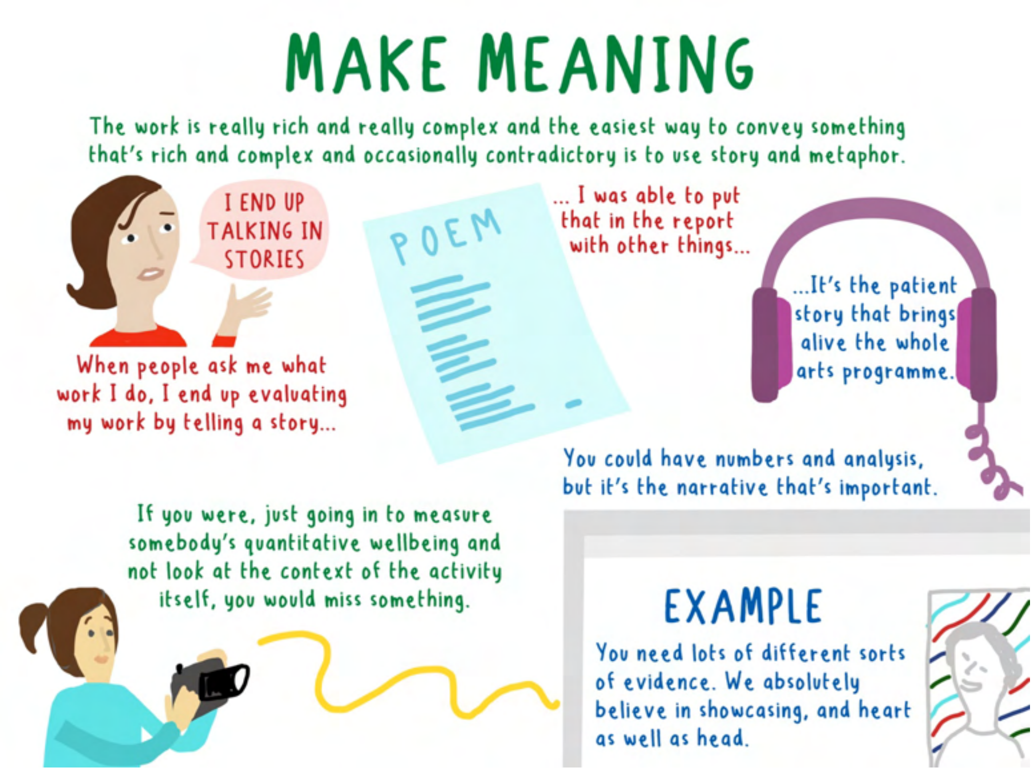
BE REFLECTIVE

If you can embed the creative process in reflective processes, you will get to a depth of knowledge that you won't reach if you just have a conversation that relies solely on words used in a logical way.



That discussion, is very important…

Evaluation is the kind of sifting through and seeing what things of meaning have come out that we can learn from and that are going to change the practice.



MAKE MEANING

The work is really rich and really complex and the easiest way to convey something that's rich and complex and occasionally contradictory is to use story and metaphor.

When people ask me what work I do, I end up evaluating my work by telling a story...

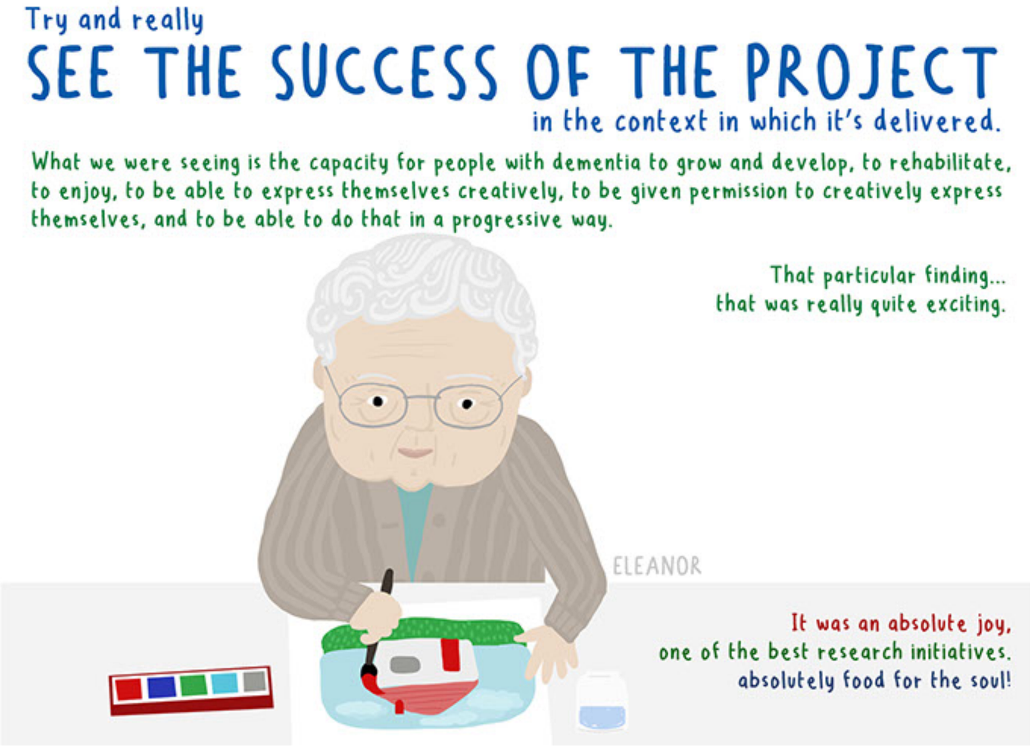
... I was able to put that in the report with other things...

...It's the patient story that brings alive the whole arts programme.

You could have numbers and analysis, but it's the narrative that's important.

If you were, just going in to measure somebody's quantitative wellbeing and not look at the context of the activity itself, you would miss something.

You need lots of different sorts of evidence. We absolutely believe in showcasing, and heart as well as head.



Try and really

SEE THE SUCCESS OF THE PROJECT

in the context in which it's delivered.

What we were seeing is the capacity for people with dementia to grow and develop, to rehabilitate, to enjoy, to be able to express themselves creatively, to be given permission to creatively express themselves, and to be able to do that in a progressive way.

That particular finding... that was really quite exciting.

It was an absolute joy, one of the best research initiatives, absolutely food for the soul!

7.4 Strengths and limitations of this study

The thesis I have written is neither the one I imagined it would be when I began, nor the one someone else would have written. Research is an active and iterative process. It is never complete. That's almost certainly a good thing, but it can leave one feeling humble in the face of everything one has not succeeded in understanding.

From a personal perspective, I have sometimes struggled to understand what disciplinary identity this PhD thesis should reflect. Its methods draw on those of the social sciences, its ambitions lie in the field of applied evaluation practice, and my approach to the subject has been informed by my previous doctoral level study of literature and the humanities. My supervisory team includes expertise from a variety of areas. The research has been delivered under the auspices of a specialist association focused on dementia sitting within a school of allied health sciences. In the end, this is the discipline that will award the degree, and thus – the discipline whose markers of research quality it must adopt. One of the guides for the methodological framework for this thesis, Charmaz, proposes that, for a constructivist grounded theory study, markers of research quality should include credibility, resonance, usefulness, and originality. The meanings of these terms are likely, however, to mean quite different things to researchers, depending on their discipline and the paradigm from which they approach their study, and therefore disciplinary, evidentiary, or aesthetic issues might also need to be taken into consideration (2014, pp. 336–337).

Charmaz further suggests that markers for credibility include evidenced familiarity with the topic, sufficiency and range of data, and a systematic approach. In response to these, I suggest that the primary strengths of this study lie in the links it makes between reported methodological challenges in the field of arts and dementia and the experiences of evaluation stakeholders. Through a literature review and hermeneutic narrative synthesis attending to both reported and unacknowledged limitations and challenges in the literature, I have identified key categories that link to and underpin the methodological challenges relating to the explanation, justification, description, and application of method. These categories form what I termed the warp thread in the fabric of methodological challenge, and became sensitising concepts used in conducting a two-phased series of in-depth stakeholder interviews. Themes identified in the first 21 interviews were checked and tested in a further five.

My decision to view the work of methodology as a social rather than a technical activity helped me to bring my prior experience and understanding of conducting evaluation in arts and health to the study. This gave me a useful perspective onto the practical difficulties facing people working outside the university setting as artists and creative practitioners, managers in arts and cultural organisations, or health and care workers involved in arts activity. Because arts and dementia practice is largely driven by the needs and structures prevalent in the world of the arts rather than that of research, I have tried to reflect the needs, opinions. and experiences of those in the arts, and those of people affected by dementia, as well as members of the academic research community who work in this field. This, along with my theoretical influences and methods, means that the study findings are firmly grounded in the direct experiences of individuals and that, hopefully, my findings might resonate with them. Throughout I have taken every opportunity offered to communicate with and disseminate my thinking to diverse communities through conference and other presentations.

The specific limitations of my literature review process, and those relating to the methods used for this research have been discussed above, in Chapters 2 and 4 respectively.

Now that it comes to discussing limitations for the study as a whole, it is challenging to know where to start, except with the general: in line with the theoretical influences on this study I can say that I reject the possibility of having any objective or certain knowledge of the world and believe that we view the world through the lens of theory, which is constructed (Maxwell, 2012). This means that I must recognise alternative accounts as both possible and desirable if the aim is to move on our understandings of any socially determined question (Charmaz, 2014) and to open up the opportunities for critical inquiry (Charmaz, 2017).

In choosing a direction for this research, I had to close some doors. Some of these, of course, were never open to me because I approached the study with a question that had already been broadly set by the parameters of a funded project, and as a result of the discipline in which I have conducted the research. My own skills and experience, and the resources available to me as a PhD student, also set boundaries around the work.

The reported difficulties of conducting research and evaluation in this field consistently centre around well-documented epistemological clashes between the disciplines involved. They often reference calls for greater methodological rigour – even if authors are rejecting such calls. If I had wished to focus on the question of how to increase rigour on either side of the epistemological divide, I would have asked quite a different question, and the answer might have spoken to a different audience. This study never aimed to provide a detailed technical analysis of the methodological difficulties entailed in conducting experimental studies or trials in this field. While this might have been highly useful to researchers, I would contend that this should, in fact, be considered a strength rather than a limitation of this work, since such an analysis might only serve to further embed some of the challenges rather than to illuminate or counter them.

Although the study draws on philosophical ideas in examining methodological challenge, it is fundamentally an exploration whose ambitions lie in applied fields rather than philosophy of methods. These ambitions were important because of understandings gained as a result of my own experience (section 4.5.1) and because most evaluation is conducted outside the academy. I originally had ambitions to identify solutions and test one or more of them in evaluation as part of the study, but, disappointingly for me, this proved ambitious given the time and resources available (section 3.4). It could be said that the study does not go far enough in either direction, providing only an outline theoretical exploration of the practical challenges and signposting solutions rather than testing their application in practice. I acknowledge the validity of this criticism, but welcome the opportunity it presents for future research, and the implications of this are discussed further in section 7.4 below.

All interviews for this research were of stakeholders involved in arts and dementia activity in the UK, and therefore its findings have been contextualised through discussion of the UK-based cultures and structures that surround them. My literature review was limited to texts in the English language, although not to projects conducted only in the UK. International policy contexts, and arts, cultural, and dementia care practices will differ widely, and therefore there are limits to the international applicability of the discussion in this thesis. Nevertheless, its central theory – the discussion of value and meaning – and its suggestion of a values-informed approach to evaluation – should have resonance wherever evaluation of arts and dementia activity is conducted and regardless of prevailing epistemic, social, or cultural values.

Perhaps the most serious criticism that could be made of this study is that while the stakeholder group for evaluation would include people directly affected by dementia who have participated in activities and their evaluation, none of these people were interviewed for the research. I could, perhaps, have chosen to seek the opinions of individual project participants, or sought the advice of a group of people with lived experience. I did not, for reasons explained above (section 4.2.1). It is somewhat ironic however that, as I have discussed above, a similar lack of representation in other studies reflects one of the key methodological challenges for the field.

7.4.1 Reflections on novel methodological aspects of this study

To frame data collection and analysis in this study, I chose to combine a critical realist-inspired approach with constructivist grounded theory (section 3.4.1). Reviewing the literature in the light of critical realist philosophy suggested to me that, beneath the obvious technical and logistical challenges for evaluation research in the field, there were likely to be underlying struggles with concepts, such as value or ethics, which might be having real effect. As a result I felt that the study would benefit from an approach to collecting and interpreting empirical (qualitative) data in which the experiences of individuals involved in evaluation could be explored in relation to such concepts, as well as to their surrounding structures and cultures. The approach I used meant that I was able to collect and analyse data in the light of pre-existing theoretical knowledge and sensitising analytical categories derived from the literature review, and to re-examine them using grounded theory's distinctive comparative approach to sampling, data collection, and coding.

Memo-writing was a useful and formative part of the analysis process, and ensured that it remained grounded in the data. Charmaz refers to memo-writing as ‘pivotal’; in my experience this feels an accurate descriptive term to use. In particular the memos I wrote as ‘connecting narratives’ allowed me to re-shape and critically reflect on the themes identified in individual interviews, and to pivot my perspective so that I could connect these themes to the body of data as a whole. So, for example, the narrative cited in full in the thesis above (section 4.4.5.1), shows how, for this particular story of evaluation practice, themes of aesthetic quality, reflective practice, value, and collaborative working are woven together.

The five interviews conducted in the second phase of interviewing were also key in questioning and validating these themes, as well as in providing material to suggest potential solutions to the challenges identified. For these I used a more realist interviewing style than in the earlier interviewers, although their topic guides were informed by my analysis from the first set of interviews. In the final analysis, my approach enabled me to suggest the centrality to the subject of evaluation stakeholders’ engagement with the concept of *value*. Further, it showed that differences between values that were *prudential* (to do with what is best for an individual) and *epistemic* (relating to the production of knowledge) can be seen as playing key component parts within their experiences of methodological challenge.

Applying a blended methodology using the two methodological elements was therefore successful, although not easy. Although there is a small body of existing literature around their combination, it is not sufficient to provide guidance on application in practice for a particular project. As an example, the method I used to create ‘connecting narratives’ was influenced by discussion around the application of realist approaches to qualitative research in the social sciences (Danermark *et al.*, 2002; Maxwell, 2012), but I had no examples to guide me in developing a template to support the creation of the narratives. Similarly, my decision to test some specific propositions derived from the first phase of interviews with ‘expert’ participants in the second phase led me to apply some recommended techniques for realist interviewing (Manzano, 2016), rather than to continue with questions to explore and elicit experience as I had done with the earlier interviews. But each methodological step and change of this kind required careful consideration and adaptation. As a result both the data collection and the analysis processes took longer than had been anticipated.

One of the benefits of PhD study is that it can provide a space in which a researcher is able to experiment with novel methodologies or combinations of methodologies. In my study, the experiences of stakeholders were central to development of my understanding of what it is about evaluation in arts and dementia that is challenging. With provisos around the time required to develop its processes, on balance, I do consider the combination of critical realism and grounded theory to have been appropriate for this study. It was not, however, an approach that produced easily identifiable mechanistic results, for example, the mapping of barriers and facilitators to evaluation practice, or outlining of practical solutions for future evaluators in particular contexts. My experience therefore confirms the suggestion that whether and how the two approaches can be successfully blended will depend on the aims and objective of the individual research project (Hoddy, 2019).

A further novel methodological aspect of my study was my work with an artist (see section 4.4.7 where I outline the details and discuss challenges and benefits). This both supported analysis and has and will enable more effective and wide-reaching dissemination of findings. Again, it should be noted that the methods and processes we used were, again, largely without precedent, required thought and reflection, and were time-consuming and resource-intensive. The artist had a particular set of skills and experience derived from her work as an animator used to working in health contexts; perhaps as a result she displayed a keen sensitivity to the interview transcript material we used. All of these elements were integral to the success of the work we did together. While this aspect of my study was both enjoyable to do and gave me a valuable perspective on my results, I could not recommend another researcher using a similar approach without ensuring sufficient time, resources, and the involvement of a collaborating artistic partner with the requisite skills.

7.5 Recommendations for research and practice

There are opportunities for translation of this research through applied projects. These would require strong partnerships outside the research community, with arts organisations, policymakers, funders, commissioners and, of course, with people with dementia. Goals of these might include the provision of training and resources for those working in the field.

In particular, future research should explore, define, test and re-define the requirements of a values-informed approach to evaluation of arts-based activities for people with dementia, perhaps aligning with work to outline the implications of values-based practice in dementia care (Hughes and Williamson, 2019) or to explore the implications for evaluation practice of applying different models of disability or enquiry into the concept of rights in the context of dementia (Shakespeare, Zeilig and Mittler, 2019).

In addition, or alongside this, research could usefully explore the value adequacy (epistemic and non-epistemic) of the key outcomes commonly associated with arts activity for people with dementia. This would entail methodological attention to the common wellbeing and quality of life self-report and observational measures used. It would also require attention to the question of whether *prudential value* – the things that most matter to the people who are targeted by arts activities – differs from the constructs that are currently measured (Taylor, 2013). The development and validation of appropriate or alternative measures, including those relating to wellbeing, might be one practical objective of such research. This would be particularly useful if they were designed to be applied easily by arts and cultural organisations and others needing to evaluate their work and could therefore contribute to a co-ordinated approach to evaluation across the sector.

Future research might also attend to the question of how and whether the arts and culture can be of value to those people who they do not currently reach. In doing this, it could have much to learn from research exploring the potential and definitions of little ‘c’ or everyday creativity (Silvia *et al.*, 2014; Camic *et al.*, 2018; Bellass *et al.*, 2019), and make important connections with research and practice in the area of ‘leisure’ more generally. It would also seek to support the work of those who provide or allocate funding for services and care for people with dementia, by ensuring that evaluation and project delivery can focus on findings that respond to these values and are confident and open about acknowledging their influence.

More practically, there is a need for development (or co-development) of consultation processes that might enable evaluation partners and stakeholders to express and understand the values that might drive their involvement, prior to the commencement of an evaluation and the project or programme delivery. Methods that might be applied to this include those derived from applied epistemology (O'Rourke and Crowley, 2013), used in projects seeking to increase the inclusion of people living with dementia (Williams *et al.*, 2020), or informed by arts-based methodologies of knowledge co-production and community engagement (Kelemen, Surman and Dikomitis, 2018).

To access the experience of artists, research could also explore how artists use models of reflective practice as everyday evaluative tools in monitoring and improving the quality of the work they do with people with dementia. It would be illuminating for a health or care researcher to work collaboratively with an artist in refining their respective understandings about the purposes and methods used for evaluation and reflection, with a view to supporting the future practice of both. Here it might be appropriate to apply to this process the understandings derived from Cousins' (2018) taxonomic exploration of the component parts of arts and dementia interventions.

7.6 In summary

The findings of this study suggest that stakeholders struggle in different ways with barriers resulting from difficulties experienced in reconciling epistemic and non-epistemic values. They also struggle to establish and define what is meaningful and what can be measured when justifying, describing, and explaining their methodological choices for evaluating arts and dementia activities. The study has included reflection on their expression of these difficulties in the light of evaluation experiences they described in interviews. To link these individual experiences to the structures that influence and surround them, I have situated my findings and reflections within a discussion of the wider fields of UK cultural policy, the prevailing paradigms for evidence and knowledge production, and in relation to funding for the arts and for dementia care and research.

As dementia diagnoses become more prevalent globally, it becomes increasingly urgent to encourage responses to the condition that will incorporate prevention, diagnosis, treatment, and care (Pickett *et al.*, 2018). The arts and culture and the energy of creative practitioners globally represent an important resource with great potential to contribute to this endeavour, particularly in relation to treatment and care. Because this is a new field, we have the opportunity to shape and influence its course globally, now. High-quality evaluation can contribute to this. This thesis advances understanding of how evaluation of arts-based activities for people with dementia is delivered. It has identified key challenges lying not only in technical details of study design, data collection and analysis, or reporting, but in compromises and clashes around how values are acknowledged and accounted for in evaluation, along with difficulties in meaning and measurement. No previous work has examined these issues in depth.

As an extended piece of research and dissemination, this project has sought to lay the ground for development of an alternative evaluation approach that could be tailored and adapted for use across the evaluation spectrum, and which has the potential to drive forward practice as well as research. This is reflected within a series of signposts for solutions for evaluators in the field. Central to these is the novel suggestion that we might want to explore the implications of adopting a values-informed approach to evaluation.

Finally, I have highlighted ‘use’ as an important, but poorly understood connecting concept. I have suggested that to unpick the fabric of methodological challenges for evaluation it may be illuminating to reconsider the use to which evaluation is put, and to do so in the context of a reframing of arts and dementia activity as an ethical practice.

8 REFERENCES

Abend, G. (2008) ‘The meaning of “Theory”’, *Sociological Theory*. Washington: American Sociological Association, 26(2), pp. 173–199.

Abendstern, M. *et al.* (2019) ‘Reflecting on the research encounter for people in the early stages of dementia : Lessons from an embedded qualitative study’, *Dementia*, (online first June 13 2019). doi: 10.1177/1471301219855295.

Alasuutari, P., Bickman, L. and Brannen, J. (eds) (2008) *The SAGE handbook of social research methods*. London: SAGE Publications Ltd.

All-Party Parliamentary Group on Dementia (2019) *Hidden no more: dementia and disability*. Available at: https://www.alzheimers.org.uk/sites/default/files/2019-06/APPG_on_Dementia_2019_report_Hidden_no_more_dementia_and_disability_media.pdf (Accessed: 14 January 2020).

All Party Parliamentary Group on Arts Health and Wellbeing (2017) *Creative health: the arts for health and wellbeing*. London. Available at: http://www.artshealthandwellbeing.org.uk/appg-inquiry/Publications/Creative_Health_Inquiry_Report_2017.pdf%0Ahttp://www.artshealthandwellbeing.org.uk/appg-inquiry/Briefings/WWCW.pdf (Accessed: 1 December 2019).

Alonso-Yanez, G. *et al.* (2019) ‘Mobilizing transdisciplinary collaborations: collective reflections on decentering academia in knowledge production’, *Global*

Sustainability, 2(e5). doi: 10.1017/sus.2019.2.

Alzheimer Europe (2011) *The ethics of dementia research*. Available at: https://www.alzheimer-europe.org/content/download/57032/363226/file/AE Report 2011 - The ethics of dementia research.pdf (Accessed: 14 January 2020).

Archer, M. *et al.* (eds) (1998) *Critical realism: essential readings*. Abingdon: Routledge.

Archer, M. S. (2007) *Making our way through the world: human reflexivity and social mobility*. Cambridge: Cambridge University Press.

Arts Council England (2013) *Great art and culture for everyone: 10-year strategic framework 2010-2020*. Manchester. Available at: http://www.artscouncil.org.uk/advice-and-guidance/browse-advice-and-guidance/great-art-and-culture-everyone (Accessed: 14 January 2020).

Ashcroft, R. E. (2004) ‘Current epistemological problems in evidence based medicine’, *Journal of Medical Ethics*, 30(2), pp. 131–135. doi: 10.1136/jme.2003.007039.

Atkinson, P. and Silverman, D. (1997) ‘Kundera’s immortality: the interview society and the invention of the self’, *Qualitative Inquiry*, 3(3), pp. 304–325. doi: 10.1177/107780049700300304.

Axelrod, R. and Cohen, M. D. (1999) *Harnessing complexity: organizational implications of a scientific frontier*. New York: Free Press.

van Baalen, A. *et al.* (2010) ‘How to evaluate quality of care from the perspective of people with dementia: an overview of the literature’, *Dementia*, 10(1), pp. 112–137.

Bach-Mortensen, A. M. and Montgomery, P. (2018) ‘What are the barriers and facilitators for third sector organisations (non-profits) to evaluate their services? A systematic review’, *Systematic Reviews*. Systematic Reviews, 7(1). doi: 10.1186/s13643-018-0681-1.

Bammer, G. (2013) *Disciplining interdisciplinarity: integration and implementation sciences for researching complex real-world problems*. Canberra: Australian National University E Press. doi: 10.26530/oapen_459901.

Barnes, V. *et al.* (2003) ‘Deconstructing dissemination: dissemination as qualitative research', *Qualitative Research*, 3(2), pp. 147–164.

Bartlett, R. (2012) 'Modifying the diary interview method to research thelives of people with dementia', *Qualitative Health Research*, 22(12), pp. 1717–1726. doi: 10.1177/1049732312462240.

Bartlett, R. (2015) 'Visualising dementia activism: using the arts to communicate research findings', *Qualitative Research*, 15(6), pp. 755–768.

Basting, A. (2014) *When art is the only medicine*, *TEDxUWMilwaukee presentation*. Available at: https://www.youtube.com/watch?v=cPA6lklMQxM (Accessed: 14 January 2020).

Batsch, N. L. and Mittelman, M. S. (2012) *World Alzheimer Report 2012: overcoming the stigma of dementia*. London. Available at: papers2://publication/uuid/889ADA69-3B6E-4DCC-BF40-936650E90072 (Accessed: 14 January 2020).

Baum, F., MacDougall, C. and Smith, D. (2006) 'Participatory action research', *Journal of Epidemiology & Community Health*, 60(10), pp. 854–857. doi: 10.1136/jech.2004.028662.

Beard, R. (2011) 'Art therapies and dementia care: a systematic review', *Dementia*, 11(5), pp. 633–656. doi: 10.1177/1471301211421090.

Belfiore, E. (2004) 'Auditing culture: the subsidised cultural sector in the New Public Management', *International Journal of Cultural Policy*, 10(2), pp. 183–202. doi: 10.1080/1028663042000255808.

Belfiore, E. and Bennett, O. (2007) 'Rethinking the social impacts of the arts', *International Journal of Cultural Policy*, 13(2), pp. 135–151. doi: 10.1080/10286630701342741.

Belfiore, E. and Bennett, O. (2010) 'Beyond the "Toolkit Approach": arts impact evaluation research and the realities of cultural policy-making', *Journal for Cultural Research*, 14(2), pp. 37–41. doi: 10.1080/14797580903481280.

Belfrage, C. and Hauf, F. (2017) 'The gentle art of retroduction: critical realism, cultural political economy and critical grounded theory', *Organization Studies*, 38(2), pp. 251–271. doi: 10.1177/0170840616663239.

Bellass, S. *et al.* (2019) 'Broadening the debate on creativity and dementia: a critical approach', *Dementia*, 18(7–8), pp. 2799–2820. doi: 10.1177/1471301218760906.

Van Belle, S. *et al.* (2016) 'Can "realist" randomised controlled trials be genuinely realist?', *Trials*. Trials, 17(1), p. 313. doi: 10.1186/s13063-016-1407-0.

Belver, M. H. *et al.* (2018) 'Art museums as a source of well-being for people with dementia: an experience in the Prado Museum', *Arts and Health*. Taylor & Francis, 10(3), pp. 213–226. doi: 10.1080/17533015.2017.1381131.

Bethell, J. *et al.* (2018) 'Patient engagement in research related to dementia: a scoping review', *Dementia*, 17(8), pp. 944–975. doi: 10.1177/1471301218789292.

Beuscher, L. and Grando, V. T. (2009) 'Challenges in conducting qualitative research with individuals with dementia', *Research in Gerontological Nursing*, 2(1), pp. 6–11. doi: 10.3928/19404921-20090101-04.

Bhaskar, R. (2008) *A realist theory of science*. 2nd ed. London: Verso.

Boell, S. K. and Cecez-Kecmanovic, D. (2014) 'A hermeneutic approach for conducting literature reviews and literature searches', *Communications of the Association for Information Systems*, 34(1), pp. 257–286. doi: 10.17705/1cais.03412.

Boland, A., Cherry, M. G. and Dickson, R. (2017) *Doing a systematic review: a student's guide*. 2nd ed. Edited by A. Boland, M. G. Cherry, and R. Dickson. London: SAGE Publications Ltd.

Booth, A. (2001) *Cochrane or cock -eyed? How should we conduct systematic reviews of qualitative research?*, *Qualitative Evidence-based Practice Conference, Taking a Critical Stance. Coventry University*. Education-line database. Available at: https://www.researchgate.net/publication/277255779%0ACochrane (Accessed: 14 January 2020).

Bourdieu, P. (1984) *Distinction: a cultural critique of the judgement of taste*. Cambridge, MA: Cambridge University Press.

Bowling, A. *et al.* (2015) 'Quality of life in dementia: a systematically conducted narrative review of dementia-specific measurement scales', *Aging & Mental Health*, 19(1), pp. 13–30. doi: 10.1080/13607863.2014.915923.

Braithwaite, J. (2017) *Turning what we know into what we do*. Available at: http://researchaustralia.org/wp-content/uploads/2017/08/Research-Australia_Turning-what-we-know-into-what-we-do_August-2017.pdf (Accessed: 2 December 2019).

Braithwaite, J. (2018) 'Changing how we think about healthcare improvement', *BMJ (Online)*, 361. doi: 10.1136/bmj.k2014.

Braun, V. and Clarke, V. (2006) 'Using thematic analysis in psychology', 3(2), pp. 77–101. doi: 10.1191/1478088706qp063oa.

Braun, V. and Clarke, V. (2014) 'What can "thematic analysis" offer health and wellbeing researchers?', *International Journal of Qualitative Studies on Health and Well-being*, 9(1), p. 26152. doi: 10.3402/qhw.v9.26152.

Braun, V. and Clarke, V. (2019) 'To saturate or not to saturate? questioning data saturation as a useful concept for thematic analysis and sample-size rationales', *Qualitative Research in Sport, Exercise and Health*, (Online first 26 December 2019). doi: 10.1080/2159676X.2019.1704846.

Breckon, J. and Dodson, J. (2016) *Using evidence: what works ? A discussion paper*. London. Available at: https://www.alliance4usefulevidence.org/assets/Alliance-Policy-Using-evidence-v4.pdf (Accessed: 14 January 2020).

Bridges, J. *et al.* (2017) 'Optimising impact and sustainability: a qualitative process evaluation of a complex intervention targeted at compassionate care', *BMJ Quality & Safety*, 26(12), pp. 970–977. doi: 10.1136/bmjqs-2017-006702.

Brinkmann, S. (2018) 'The interview', in Denzin, N. K. and Lincoln, Y. S. (eds) *The SAGE handbook of qualitative research*. 5th edn. Thousand Oaks, California: Sage Publications Inc., pp. 576–599.

Broderick, S. (2011) 'Arts practices in unreasonable doubt? reflections on understandings of arts practices in healthcare contexts', *Arts & Health*, 3(2), pp. 95–109. doi: 10.1080/17533015.2010.551716.

Brooker, D. and Latham, I. (2016) *Person-centred dementia care: making services better with the VIPS framework*. 2nd edn. London: Jessica Kingsley Publishers.

Brooks, J., Savitch, N. and Gridley, K. (2016) 'Removing the "gag": involving people with dementia in research as advisers and participants’, *Social Research Practice*, (3), pp. 3–14.

Brown, A. P. (2010) ‘Qualitative method and compromise in applied social research’, *Qualitative Research*, 10(2), pp. 229–248. doi: 10.1177/1468794109356743.

Bryden, C. (2020) ‘Challenging the discourses of loss: A continuing sense of self within the lived experience of dementia’, *Dementia*, 19(1), pp. 74–82. doi: 10.1177/1471301219876711.

Bungay, H. *et al.* (2014) *The value of the arts in therapeutic and clinical interventions: a critical review of the literature*. Cambridge: Anglia Ruskin University.

Camic, P. M. *et al.* (2018) ‘Conceptualising and understanding artistic creativity in the dementias: interdisciplinary approaches to research and practise on behalf of the Created Out of Mind team’, *Frontiers in Psychology*, 9, p. 1842. doi: 10.3389/fpsyg.2018.01842.

Camic, P. M., Tischler, V. and Pearman, C. H. (2014) ‘Viewing and making art together: a multi-session art-gallery-based intervention for people with dementia and their carers’, *Aging & Mental Health*, 18(2), pp. 161–168. doi: 10.1080/13607863.2013.818101.

Camic, P. M., Zeilig, H. and Crutch, S. J. (2018) ‘The arts and dementia: emerging directions for theory, research and practice’, *Dementia*, 17(6), pp. 641–644. doi: 10.1177/1471301218772972.

Carnwath, J. D. and Brown, A. S. (2014) *Understanding the value and impacts of cultural experiences: a literature review*. London: Arts Council England. doi: 10.1080/09548963.2014.967012.

CASP Critical Appraisal Skills Programme (2018) *CASP Checcklist: 10 questions to help you make sense of qualitative research*. Oxford: CASP Critical Appraisal Skills Programme. Available at: https://casp-uk.net/casp-tools-checklists/ (Accessed: 14 January 2020).

Cerejeira, J., Lagarto, L. and Mukaetova-Ladinska, E. B. (2012) ‘Behavioral and psychological symptoms of dementia’, *Frontiers in Neurology*, 3. doi:

10.3389/fneur.2012.00073.

Charlesworth, A. *et al.* (2018) *Securing the future: funding health and social care to the 2030s*. London: The Institute for Fiscal Studies.

Charmaz, K. (2014) *Constructing grounded theory*. 2nd edn. London: SAGE Publications Inc.

Charmaz, K. (2017) 'The power of constructivist grounded theory for critical inquiry', *Qualitative Inquiry*, 23(1), pp. 34–45. doi: 10.1177/1077800416657105.

Chen, H.-T. (1996) 'A comprehensive typology for program evaluation', *American Journal of Evaluation*, 17(2), pp. 121–130. doi: 10.1177/109821409601700204.

Clare, A. *et al.* (2019) 'Using music to develop a multisensory communicative environment for people with late-stage dementia', *The Psychologist*, (Online first 08 December 2019). doi: 10.1093/geront/gnz169.

Clarke, A. and Dawson, R. (1999) *Evaluation research: an introduction to principles, methods and practice*. London: SAGE Publications Ltd.

Clift, S. (2012) 'Creative arts as a public health resource: moving from practice-based research to evidence-based practice', *Perspectives in Public Health*, 132(3), pp. 120–127.

Cohen-Mansfield, J. (2018) 'Non-pharmacological interventions for persons with dementia: what are they and how should they be studied?', *International Psychogeriatrics*, 30(3), pp. 281–283. doi: 10.1017/S104161021800039X.

Collier, A. (1994) *Critical realism: an introduction to Roy Bhaskar's philosophy*. London: Verso.

Cook, T. D. and Wong, V. C. (2008) 'Better quasi-experimental practice', in Alasuutari, P., Bickman, L., and Brannen, J. (eds) *The SAGE handbook of social research methods*. London: SAGE Publications Ltd., pp. 134–165.

Corley, K. G. (2015) 'A commentary on "What Grounded Theory Is…": engaging a phenomenon from the perspective of those living it', *Organizational Research Methods*, 18(4), pp. 600–605. doi: 10.1177/1094428115574747.

Cousins, E. (2018) *Making meaning, creating connection : A taxonomy of arts* *interventions for people with dementia*. Unpublished PhD thesis. University of Nottingham.

Cousins, E. *et al.* (2019a) ‘A taxonomy of arts interventions for people with dementia’, *The Gerontologist*, 60(1), pp. 124–134. doi: 10.1093/geront/gnz024.

Cousins, E. *et al.* (2019b) ‘Principles and features to define and describe arts interventions for people with dementia: a qualitative realist study’, *Arts and Health*, 11(3), pp. 202–218. doi: doi.org/10.1080/17533015.2018.1490787.

Cowl, A. L. and Gaugler, J. E. (2014) ‘Efficacy of creative arts therapy in treatment of Alzheimer’s disease and dementia: a systematic literature review’, *Activities, Adaptation & Aging*, 38(4), pp. 281–330. doi: 10.1080/01924788.2014.966547.

Craig, P. *et al.* (2008) *Developing and evaluating complex interventions: new guidance*, *Medical Research Council*. London: Medical Research Council. Available at: www.mrc.ac.uk/complexinterventionsguidance.

Cridland, E. K. *et al.* (2016) ‘Reflections and recommendations for conducting in-depth interviews with people with dementia’, *Qualitative Health Research*, 26(13), pp. 1774–1786.

Crossick, G. and Kaszynska, P. (2016) *Understanding the value of arts & culture: the AHRC Cultural Value project*. Swindon: Arts & Humanities Research Council.

Crotty, M. (1998) *The foundations of social research: meaning and perspective in the research process*. London: Sage Publications, Ltd.

Cummings, J. and Bennett, V. (2012) *Compassion in practice: nursing, midwifery and care staff - our vision and strategy*. London: NHS Commissioning Board, Department of Health.

Cutler, D. and Bakewell, D. J. (2009) *Ageing artfully: older people and professional participatory arts in the UK*. London: The Baring Foundation.

Damschroder, L. J. *et al.* (2009) ‘Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science’, *Implementation Science*, 4(1), p. 50. doi: 10.1186/1748-5908-4-50.

Danermark, B. *et al.* (2002) *Explaining society: critical realism in the social sciences*. London: Routledge.

Davies, M. and Heath, C. (2013) *Evaluating evaluation: increasing the impact of summative evaluation in museums and galleries*. London: King's College London.

Daykin, N. *et al.* (2016) *Creative and credible: developing evaluation resources for arts, health and wellbeing.* Bristol: University of the West of England.

Daykin, N. *et al.* (2017) 'Creative and credible evaluation for arts, health and well-being: opportunities and challenges of co-production', *Arts & Health*, 9(2), pp. 123–138. doi: 10.1080/17533015.2016.1206948.

Daykin, N. (2020) *Arts, health and wellbeing: A critical perspective on research, policy and practice*. Abingdon: Routledge.

Daykin, N., Attwood, M. and Willis, J. (2013) 'Supporting arts and health evaluation: report of a UK knowledge transfer partnership', *Journal of Applied Arts & Health*, 4(2), pp. 179–190.

Daykin, N. and Joss, T. (2016) *Arts for health and wellbeing: an evaluation framework*. London: Public Health England.

Decoteau, C. L. (2016) 'The AART of ethnography: a critical realist explanatory research model', *Journal for the Theory of Social Behaviour*, 47(1). doi: 10.1111/jtsb.12107.

DeNora, T. and Ansdell, G. (2014) 'What can't music do?', *Psychology of Well-Being: Theory, Research and Practice*, 4(23).

Denzin, N. K. and Lincoln, Y. S. (2018) *The SAGE handbook of qualitative research*. 5th edn. Edited by N. K. Denzin and Y. S. Lincoln. Thousand Oaks, California: SAGE Publications Ltd.

DeVito, N. J. and Goldacre, B. (2019) 'Catalogue of bias: publication bias', *BMJ Evidence-Based Medicine*, 24(2), pp. 53–54. doi: 10.1136/bmjebm-2018-111107.

Dewey, J. (1934) *Art as experience*. Reprint. New York: Penguin (Perigree trade paperback 2005).

Dewing, J. (2008) 'Process consent and research with older persons living with dementia', *Research Ethics*, 4(2), pp. 59–64. doi: 10.1177/174701610800400205.

Dixon-Woods, M. *et al.* (2006) ‘Conducting a critical interpretive synthesis of the literature on access to healthcare by vulnerable groups’, *BMC Medical Research Methodology*, 6, pp. 1–13. doi: 10.1186/1471-2288-6-35.

Dixon-Woods, M. (2006) ‘How can systematic reviews incorporate qualitative research? a critical perspective’, *Qualitative Research*, 6(1), pp. 27–44.

Dixon-Woods, M., Booth, A. and Sutton, A. J. (2007) ‘Synthesizing qualitative research: A review of published reports’, *Qualitative Research*, 7(3), pp. 375–421. doi: 10.1177/1468794107078517.

Dowlen, R. *et al.* (2018) ‘The personal benefits of musicking for people living with dementia: a thematic synthesis of the qualitative literature’, *Arts & Health*, 10(3), pp. 197–212. doi: 10.1080/17533015.2017.1370718.

Dowson, B., McDermott, O. and Schneider, J. (2019) ‘What indicators have been used to evaluate the impact of music on the health and wellbeing of people with dementia? a review using meta-narrative methods’, *Maturitas*, 127, pp. 26–34. doi: 10.1016/j.maturitas.2019.06.001.

Dupuis, S. L. *et al.* (2016) ‘Re-claiming citizenship through the arts’, *Dementia*, 15(3), pp. 358–380. doi: 10.1177/1471301216637206.

Eakin, J. M. and Mykhalovskiy, E. (2003) ‘Reframing the evaluation of qualitative health research: reflections on a review of appraisal guidelines in the health sciences’, *Journal of Evaluation in Clinical Practice*, 9(2), pp. 187–194. doi: 10.1046/j.1365-2753.2003.00392.x.

Eekelaar, C., Camic, P. M. and Springham, N. (2012) ‘Art galleries, episodic memory and verbal fluency in dementia: an exploratory study’, *Psychology of Aesthetics, Creativity, and the Arts*, 6(3), pp. 262–272. doi: 10.1037/a0027499.

Egger, M., Smith, G. D. and O’Rourke, K. (2001) ‘Rationale, potentials, and promise of systematic reviews’, in *Systematic Reviews in Health Care: Meta-analysis in Context*. 2nd edn. London: BMJ Publishing Group, pp. 3–19.

Eigenbrode, S. D. *et al.* (2007) ‘Employing philosophical dialogue in collaborative science’, *BioScience*, 57(1), pp. 55–64. doi: 10.1641/B570109.

Elliott, K. C. (2017) *A tapestry of values: an introduction to values in science*. Oxford: Oxford University Press.

Elliott, K. C. and McKaughan, D. J. (2014) ‘Nonepistemic values and the multiple goals of science’, *Philosophy of Science*, 81(1), pp. 1–21. doi: 10.1086/674345.

Erickson, F. (2017) ‘A history of qualitative inquiry in social and educational research’, in Denzin, N. K. and Lincoln, Y. S. (eds) *The SAGE handbook of qualitative research*. 5th ed. Thousand Oaks, California: SAGE Publications Ltd, pp. 36–65.

Fancourt, D. (2017) *Arts in health: Designing and researching interventions*. Oxford: Oxford University Press.

Fancourt, D. and Finn, S. (2019) *What is the evidence on the role of the arts in improving health and well-being? A scoping review*. Available at: http://www.euro.who.int/en/publications/abstracts/what-is-the-evidence-on-the-role-of-the-arts-in-improving-health-and-well-being-a-scoping-review-2019.

Fancourt, D. and Joss, T. (2015) ‘Aesop: A framework for developing and researching arts in health programmes’, *Arts & Health*. Routledge, 7(1), pp. 1–13. doi: 10.1080/17533015.2014.924974.

Fancourt, D., Steptoe, A. and Cadar, D. (2018) ‘Cultural engagement and cognitive reserve : museum attendance and dementia incidence over a 10-year period’, *The British Journal of Psychiatry*, pp. 13–15. doi: 10.1192/bjp.2018.129.

Fletcher, A. J. (2017) ‘Applying critical realism in qualitative research: methodology meets method’, *International Journal of Social Research Methodology*. Routledge, 20(2), pp. 181–194. doi: 10.1080/13645579.2016.1144401.

Flyvbjerg, B. (2011) ‘Case study’, in Denzin, N. K. and Lincoln, Y. S. (eds) *The SAGE handbook of qualitative research*. 4th edn. Thousand Oaks, California: SAGE Publications Inc, pp. 301–316. doi: 10.1016/S1360-8592(98)80013-2.

Francis, J. J. *et al.* (2010) ‘What is an adequate sample size? Operationalising data saturation for theory-based interview studies’, *Psychology and Health*, 25(10), pp. 1229–1245. doi: 10.1080/08870440903194015.

Fujiwara, D., Kudrna, L. and Dolan, P. (2014) *Quantifying and valuing the wellbeing impacts of culture and sport*. London: Department for Culture, Media & Sport.

Fujiwara, D. and Lawton, R. (2015) *Evaluation of the Reminiscence Arts and Dementia: impact on quality of life (RADIQL) programme in six care homes*. London: Simetrica. Available at: http://www.age-exchange.org.uk/what-we-do/inspired-caring/radiql/.

Fulford, K. W. M. (2011) 'The value of evidence and evidence of values: bringing together values-based and evidence-based practice in policy and service development in mental health', *Journal of Evaluation in Clinical Practice*, 17(5), pp. 976–987. doi: 10.1111/j.1365-2753.2011.01732.x.

Galloway, S. (2009) 'Theory-based evaluation and the social impact of the arts', *Cultural Trends*, 18(2), pp. 125–148. doi: 10.1080/09548960902826143.

Garabedian, C. E. and Kelly, F. (2018) 'Haven: sharing receptive music listening to foster connections and wellbeing for people with dementia who are nearing the end of life, and those who care for them', *Dementia*, (Online first 11 October 2018). doi: 10.1177/1471301218804728.

Gartlehner, G. *et al.* (2006) 'A simple and valid tool distinguished efficacy from effectiveness studies', *Journal of Clinical Epidemiology*, 59(10), pp. 1040–1048. doi: 10.1016/j.jclinepi.2006.01.011.

Gibson, R. and Robichaud, S. (2017) 'Evaluating Dancing With Parkinson's: reflections from the perspective of a community organization', *Evaluation and Program Planning*. Elsevier, (May). doi: 10.1016/j.evalprogplan.2017.05.010.

Glaser, B. G. (1978) *Theoretical sensitivity*. Mill Valley, CA: The Sociology Press.

Glaser, B. G. (1992) *The basics of grounded theory analysis: emergence vs forcing*. Mill Valley, CA: Sociology Press.

Glaser, B. G. (1999) 'The future of grounded theory', *Qualitative Health Research*, 9(6), pp. 836–845. doi: 10.1177/104973299129122199.

Glaser, B. G. (2001) *The grounded theory perspective: COnceptualization contrasted with description*. Mill Valley, CA: Sociology Press.

Glaser, B. and Strauss, A. (1967) *The discovery of grounded theory: strategies for qualitative research*. New Brunswick: Transaction Publishers.

Godin, K. *et al.* (2015) 'Applying systematic review search methods to the grey literature: a case study examining guidelines for school-based breakfast programs in Canada', *Systematic Reviews*. Systematic Reviews, 4(1), p. 138. doi: 10.1186/s13643-015-0125-0.

Goldenberg, M. J. (2006) 'On evidence and evidence-based medicine: lessons from the philosophy of science', *Social Science & Medicine*, 62(11), pp. 2621–2632. doi: 10.1016/j.socscimed.2005.11.031.

Goldman, R. (2006) 'Video representations and the perspectivity framework: epistemology, ethnography, evaluation and ethics', in Goldman, R. et al. (eds) *Video Research in the Learning Sciences*. New York, New York: Routledge, pp. 1–38.

Good, A., Lamont, E. and Triangle Consulting Social Enterprise Ltd (2018) *Outcomes Star TM Psychometric factsheet: Older Person's Star TM*. Triangle Consulting Social Enterprise Ltd. Available at: https://www.outcomesstar.org.uk/using-the-star/see-the-stars/older-persons-star/ (Accessed: 14 January 2020).

Goodman, C. *et al.* (2011) 'Culture, consent, costs and care homes: enabling older people with dementia to participate in research', *Aging & Mental Health Mental Health*, 15(4), pp. 475–481. doi: 10.1080/13607863.2010.543659.

Goulding, A. (2014) 'Arts on prescription for older people: different stakeholder perspectives on the challenges of providing evidence of impact on health outcomes', *Journal of Applied Arts & Health*, 5(1), pp. 83–107. doi: 10.1386/jaah.5.1.83_1.

Grant, M. J. and Booth, A. (2009) 'A typology of reviews: an analysis of 14 review types and associated methodologies', *Health Information and Libraries Journal*, 26(2), pp. 91–108. doi: 10.1111/j.1471-1842.2009.00848.x.

Gray, K. *et al.* (2018) 'Critical reflections on methodological challenge in arts and dementia evaluation and research', *Dementia*, 17(6), pp. 775–784. doi: 10.1177/1471301217734478.

Green, M. J. and Myers, K. R. (2010) 'Graphic medicine: use of comics in medical education and patient care', *BMJ*, 340, p. c863. doi: 10.1136/bmj.c863.

Greenhalgh, T. *et al.* (2004) 'Diffusion of innovations in service organizations:

systematic review and recommendations', *The Milbank Quarterly*, 82(4), pp. 581–629. doi: 10.1111/j.0887-378X.2005.340_1.x.

Greenhalgh, T. *et al.* (2016) 'Achieving research impact through co-creation in community-based health services: literature review and case study', *The Milbank Quarterly*, 94(2), pp. 392–429.

Greenhalgh, T. (2016) *Cultural contexts of health: the use of narrative research in the health sector*, *Health Evidence Network Report 49*. Copenhagen. Available at: http://www.euro.who.int/__data/assets/pdf_file/0004/317623/HEN-synthesis-report-49.pdf?ua=1.

Greenhalgh, T. (2017) *How to implement evidence-based healthcare*. New York: John Wiley & Sons.

Greenhalgh, T., Howick, J. and Maskrey, N. (2014) 'Evidence-based medicine: a movement in crisis?', *BMJ*, 348. doi: 10.1136/bmj.g3725.

Greenhalgh, T., Thorne, S. and Malterud, K. (2018) 'Time to challenge the spurious hierarchy of systematic over narrative reviews?', *European Journal of Clinical Investigation*, (March), p. e12931. doi: 10.1111/eci.12931.

Gregory, H. *et al.* (2012) 'Poetry in dementia care: overcoming the challenges', *Journal of Dementia Care*. Hawker Publications, 20(2), pp. 20-23 4p.

Guba, E. S. and Lincoln, Y. S. (1989) *Fourth generation evaluation*. Newbury Park, CA: SAGE Publications Inc.

Guston, D. (2001) 'Boundary organizations in environmental policy and science: an introduction', *Science, Technology, & Human Values*, 26(4), pp. 399–408.

van Haeften-Van Dijk, A. M., van Weert, J. C. M. and Droës, R. M. (2015) 'Implementing living room theatre activities for people with dementia on nursing home wards: a process evaluation study', *Aging & Mental Health*, 19(6), pp. 536–547. doi: 10.1080/13607863.2014.955459.

Hallingberg, B. *et al.* (2018) 'Exploratory studies to decide whether and how to proceed with full-scale evaluations of public health interventions: a systematic review of guidance', *Pilot and Feasibility Studies*, 4(104), pp. 1–12. doi: 10.1186/s40814-018-0290-8.

Hammersley, M. (2008) 'Assessing validity in social research', in Alasuutari, P., Bickman, L., and Brannen, J. (eds) *The SAGE handbook of social research methods*. London: SAGE Publications Ltd., pp. 42–53.

Hancock, G. R. (2004) ‘Experimental, quasi-experimental, and nonexperimental design and analysis with latent variables’, in Kaplan, D. (ed.) *The SAGE handbook of quantitative methodology for the social sciences*. Thousand Oaks, California: SAGE Publications Inc, pp. 317–334.

Hannemann, B. T. (2006) ‘Creativity with dementia patients: can creativity and art stimulate dementia patients positively?’, *Gerontology*, 52(1), pp. 59–65.

Hara, M. (2011) ‘Music in dementia care: increased understanding through mixed research methods’, *Music and Arts in Action*, 3(2), pp. 34–58.

Hawe, P., Shiell, A. and Riley, T. (2009) ‘Theorising interventions as events in systems’, *American Journal of Community Psychology*, 43(3–4), pp. 267–276. doi: 10.1007/s10464-009-9229-9.

Hellstrom, I. *et al.* (2007) ‘Ethical and methodological issues in interviewing persons with dementia’, *Nursing Ethics*, 14(5), pp. 608–619.

Hennink, M. M., Kaiser, B. N. and Marconi, V. C. (2016) ‘Code saturation versus meaning saturation: how many interviews are enough?’, *Qualitative Health Research*, 27(4), pp. 591–608. doi: 10.1177/1049732316665344.

Higgins, J. *et al.* (2019) *Cochrane handbook for systematic reviews of interventions version 6.0*. Cochrane.

Hoddy, E. T. (2019) ‘Critical realism in empirical research: employing techniques from grounded theory methodology’, *International Journal of Social Research Methodology*, 22(1), pp. 111–124. doi: 10.1080/13645579.2018.1503400.

Holden, J. (2004) *Capturing cultural value: how culture has become a tool of government policy*. London: Demos.

Holt, A. (2010) ‘Using the telephone for narrative interviewing: a research note’, *Qualitative Research*, 10(1), pp. 113–121. doi: 10.1177/1468794109348686.

Holt, N. J. (2018) ‘Using the experience-sampling method to examine the psychological mechanisms by which participatory art improves wellbeing’, *Perspectives in Public Health*, 138(1), pp. 55–65. doi: 10.1177/1757913917739041.

Hughes, J. C. (2014) ‘The aesthetic approach to people with dementia’, *International Psychogeriatrics*, 26(09), pp. 1407–1413. doi: 10.1017/S1041610214001100.

Hughes, J. C. and Williamson, T. (2019) *The dementia manifesto: putting values-based practice to work*. Cambridge, UK: Cambridge University Press.

Husserl, E. (1970) *The crisis of the European sciences and Transcendental phenomenology: an introduction to phenomenological philosophy*. Edited by D. Carr. Evanston: Northwestern University Press.

Institute of Medicine (2001) *Crossing the quality chasm: a new health system for the 21st century*. Washington: National Academy Press. doi: 10.17226/10027.

Irvine, A. (2010) *Using phone interviews*, *Realities Toolkit*. 14. Manchester. Available at: http://eprints.ncrm.ac.uk/1576/1/14-toolkit-phone-interviews.pdf.

Jensen, E. (2014) *Questioning the specious quantification of cultural impacts*, *Cultural Value Initiative*. Available at: www.culturalvalueinitiative.org (Accessed: 2 December 2019).

Jerrim, J. and De Vries, R. (2017) ‘The limitations of quantitative social science for informing public policy’, *Evidence and Policy*, 13(1), pp. 117–133. doi: 10.1332/174426415X14431000856662.

Johnson, J. *et al.* (2017) ‘Museum activities in dementia care: Using visual analog scales to measure subjective wellbeing’, *Dementia*, 16(5), pp. 591–610. doi: 10.1177/1471301215611763.

Jones, E. M. *et al.* (2015) ‘The development of narrative practices in medicine c.1960-c.2000’, in *Wellcome Witnesses to Contemporary Medicine*. London: Queen Mary, University of London.

Kahneman, D. (2011) *Thinking fast and slow*. London: Allen Lane.

Kaplan, A. (1964) *The conduct of inquiry: methodology for behavioral science*. San Francisco: Chandler Publishing Company. Available at: file://catalog.hathitrust.org/Record/000010182.

Kara, H. (2015) *Creative research methods in the social sciences: a practical guide*. Bristol: Policy Press.

Karkou, V. and Meekums, B. (2017) ‘Dance movement therapy for dementia’, *Cochrane Database of Systematic Reviews*, (2). doi: 10.1002/14651858.CD011022.pub2.

Kelemen, M., Surman, E. and Dikomitis, L. (2018) ‘Cultural animation in health research: an innovative methodology for patient and public involvement and engagement’, *Health Expectations*, 21(4), pp. 805–813. doi: 10.1111/hex.12677.

Kelly, M. P. *et al.* (2015) ‘The importance of values in evidence-based medicine’, *BMC Medical Ethics*. BMC Medical Ethics, 16(1), p. 69. doi: 10.1186/s12910-015-0063-3.

Kempster, S. and Parry, K. W. (2011) ‘Grounded theory and leadership research: a critical realist perspective’, *The Leadership Quarterly*. Elsevier B.V., 22(1), pp. 106–120. doi: 10.1016/j.leaqua.2010.12.010.

Keuth, H. (2015) ‘The positivist dispute in German sociology: A scientific or a political controversy?’, *Journal of Classical Sociology*, 15(2), pp. 154–169. doi: 10.1177/1468795X14567825.

Kinsella, E. A. (2010) ‘The art of reflective practice in health and social care: Reflections on the legacy of Donald Schön’, *Reflective Practice*, 11(4), pp. 565–575. doi: 10.1080/14623943.2010.506260.

Kitwood, T. (1997) *Dementia reconsidered: the person comes first*. Maidenhead: Open University Press.

Knight, A. D. *et al.* (2017) *A whole new world: funding and commissioning in complexity*. London: Collaborate CIC. Available at: http://doi.wiley.com/10.1111/jan.12815.

Kontos, P. C. and Poland, B. D. (2009) ‘Mapping new theoretical and methodological terrain for knowledge translation: contributions from critical realism and the arts’, *Implementation Science*, 4(1). doi: 10.1186/1748-5908-4-1.

Kuhn, T. S. (2012) *The structure of scientific revolutions*. 4th edn. Chicago, IL: University of Chicago Press.

Kvale, S. (1995) ‘The social construction of validity’, *Qualitative Inquiry*, 1(1), pp. 19–40. doi: 10.1177/107780049500100103.

Lather, P. (1992) ‘Critical frames in educational research: feminist and post-structural perspectives', *Theory Into Practice*, pp. 87–99. doi: 10.1080/00405849209543529.

Latour, B. and Woolgar, S. (1979) *Laboratory life: the construction of scientific facts*. Princeton, NJ: Princeton University Press.

Law, J. (2004) *After method: mess in social science research*. Oxford: Routledge.

Law, J. and Urry, J. (2004) 'Enacting the social', *Economy and Society*, 33(3), pp. 390–410. Available at: http://oro.open.ac.uk/21643/.

Lélé, S. and Norgaard, R. B. (2005) 'Practicing interdisciplinarity', *BioScience*, 55(11), pp. 967–975. doi: 10.1641/0006-3568(2005)055[0967:pi]2.0.co;2.

Lincoln, Y. S. and Guba, E. G. (1985) *Naturalistic inquiry*. Beverley Hills, CA: Sage Publications Inc.

Lincoln, Y. S. and Guba, E. S. (2007) 'But is it rigorous? trustworthiness and authenticity in naturalistic evaluation', *New Directions for Evaluation*, 114.

Lincoln, Y. S., Lynham, S. A. and Guba, E. G. (2017) 'Paradigmatic controversies, contradictions, and emerging confluences, revisited', in Denzin, N. K. and Lincoln, Y. S. (eds) *The SAGE handbook of qualitative research*. 5th edn. Thousand Oaks, CA: SAGE Publications Ltd, pp. 108–150.

Lipscomb, M. (2014) *A hospice in change: applied social realist theory*. Abingdon: Routledge.

Lo, C. O. (2014) 'Enhancing groundedness in realist grounded theory research', *Qualitative Psychology*, 1(1), pp. 61–76. doi: 10.1037/qup0000001.

Lorde, A. (1986) *Sister outsider*. Kindle edn. Penguin.

Ludwin, K. and Capstick, A. (2015) 'Using participatory video to understand diversity among people with dementia in long-term care', *Journal of Psychological Issues in Organizational Culture*, 5(4), pp. 30–38. doi: 10.1002/jpoc.21161.

Manzano, A. (2016) 'The craft of interviewing in realist evaluation', *Evaluation*, 22(3), pp. 342–360. doi: 10.1177/1356389016638615.

Martyr, A. *et al.* (2018) 'Living well with dementia: a systematic review and correlational meta-analysis of factors associated with quality of life, well-being and life satisfaction in people with dementia', *Psychological Medicine*, 48(13), pp. 2130–2139. doi: 10.1017/S0033291718000405.

Matarasso, F. (1996) *Defining values: evaluating arts programmes*, *The Social Impact of the Arts*. Stroud, Comedia. Available at: http://www.culturenet.cz/res/data/004/000566.pdf.

Matarasso, F. (2000) *Weighing poetry: exploring funding criteria for assessing artistic quality*. Available at: http://www.artscouncil.ie/Publications/Literature--English-language-/Weighing-poetry_-Exploring-funding-criteria-for-assessing-artistic-quality/ (Accessed: 14 January 2020).

Matarasso, F. (2013) 'Creative progression: reflections on quality in participatory arts', *UNESCO Observatory Multi-Disciplinary Journal in the Arts*, 3(3).

Mauthner, N. S. and Doucet, A. (2003) 'Reflexive accounts and accounts of reflexivity in qualitative data analysis', *Sociology*, 37(3), pp. 413–431.

Maxwell, J. A. (2008) 'The value of a realist understanding of causality for qualitative research', in Denzin, N. K. and Giardina, M. D. (eds) *Qualitative Inquiry and the Politics of Evidence*. Walnut Creek, CA: Left Coast Press, pp. 163–181.

Maxwell, J. A. (2012) *A realist approach for qualitative research*. London: SAGE Publications Ltd.

May, C. R. *et al.* (2011) 'Evaluating complex interventions and health technologies using normalization process theory: development of a simplified approach and web-enabled toolkit', *BMC Health Services Research*, 11(1), p. 245. doi: 10.1186/1472-6963-11-245.

May, C. R., Johnson, M. and Finch, T. (2016) 'Implementation, context and complexity', *Implementation Science*, 11(1), p. 141. doi: 10.1186/s13012-016-0506-3.

Mayer, R. E. and Sims, V. K. (1994) 'For whom is a picture worth a thousand words? extensions of a dual-coding theory of multimedia learning', *Journal of Educational Psychology*, 86(3), pp. 389–401. doi: 10.1037/0022-0663.86.3.389.

McCormack, B. (2003) 'Researching nursing practice: does person-centredness matter?', *Nursing*, 4, pp. 179–188.

McDermott, O. *et al.* (2013) 'Music therapy in dementia: a narrative synthesis systematic review', *International Journal Of Geriatric Psychiatry*, 28(8), pp. 781–794. doi: 10.1002/gps.3895.

de Medeiros, K. and Basting, A. (2014) '"Shall I compare thee to a dose of donepezil?": cultural arts interventions in dementia care research', *The Gerontologist*, 54(3), pp. 344–53. doi: 10.1093/geront/gnt055.

Melville, R. (2017) *The logic of evaluation in the arts: exploring artists' responses to measurement within a publicly funded arts organisation*.

Merli, P. (2002) 'Evaluating the social impact of participation in arts activities', *International Journal of Cultural Policy*, 8(1), pp. 107–118. doi: 10.1080/10286630290032477.

Midgley, G. (2003) 'Science as systemic intervention: some implications of systems thinking and complexity for the philosophy of science', *Systemic Practice and Action Research*, 16(2), pp. 77–97. doi: 10.1023/A:1022833409353.

Miller, T. R. *et al.* (2008) 'Epistemological pluralism: reorganizing interdisciplinary research', *Ecology and Society*, 13(2), p. 46.

Moore, L. *et al.* (2018) 'Exploratory studies to inform full-scale evaluations of complex public health interventions: the need for guidance', *Journal of Epidemiology and Community Health*, (Online issue publication 11 September 2018), pp. 1–2. doi: 10.1136/jech-2017-210414.

Morse, J. M. (2006) 'The Politics of Evidence', *Qualitative Health Research*, 16(3), pp. 395–404. doi: 10.1177/1049732305285482.

Murphy, K. *et al.* (2014) 'Articulating the strategies for maximising the inclusion of people with dementia in qualitative research studies', *Dementia*, 14(6), pp. 800–824. doi: 10.1177/1471301213512489.

Myerscough, J. (1989) *The economic importance of the arts in Britain*. London: Policy Studies Institute.

National Institute for Health and Clinical Excellence (2008) 'Social value judgements: Principles for the development of NICE guidance (Second edition)'. London: National Institute for Health and Clinical Excellence. doi: 10.1097/01.qmh.0000281065.68473.e7.

National Institute for Health and Clinical Excellence (2015) *Dementia - assessment, management and support for people living with dementia and their carers: guidance and guidelines*, *www.nice.org.uk*. Available at: https://www.nice.org.uk/guidance/indevelopment/gid-cgwave0792 (Accessed: 14 January 2020).

Newman, A. *et al.* (2016) 'Carrying out research across the arts and humanities and social sciences: developing the methodology for Dementia and Imagination', *Cultural Trends*, 25(4), pp. 218–232. doi: 10.1080/09548963.2016.1241338 To.

Newman, A. *et al.* (2019) 'The role of the visual arts in the resilience of people living with dementia in care homes', *Ageing and Society*, 39(11), pp. 2465–2482. doi: 10.1017/S0144686X18000594.

Noble, H. and Smith, J. (2015) 'Issues of validity and reliability in qualitative research', *Evidence Based Nursing*, 18(2), pp. 34 LP – 35. doi: 10.1136/eb-2015-102054.

Noice, T., Noice, H. and Kramer, A. F. (2013) 'Participatory arts for older adults: a review of benefits and challenges', *The Gerontologist*, 54(5), pp. 741–753. doi: 10.1093/geront/gnt138.

Novick, G. (2008) 'Is there a bias against telephone interviews in qualitative research?', *Research in Nursing & Health*, 31(4), pp. 391–398. doi: 10.1002/nur.20259.

Nowotny, H., Scott, P. and Gibbons, M. (2003) 'Mode 2 revisited: the new production of knowledge', *Minerva*, 41, pp. 179–194.

Nyman, S. R. and Szymczynska, P. (2016) 'Meaningful activities for improving the wellbeing of people with dementia: beyond mere pleasure to meeting fundamental psychological needs', *Perspectives in Public Health*, 136(2), pp. 99–107. doi: 10.1177/1757913915626193.

O'Brien, D. (2010) *Measuring the value of culture: a report to the Department for Culture Media and Sport*. London: Department for Culture, Media & Sport. doi: 10.4467/20843968ZP.12.018.0485.

O'Reilly, M. and Parker, N. (2013) '"Unsatisfactory saturation": a critical exploration of the notion of saturated sample sizes in qualitative research', *Qualitative Research*, 13(2), pp. 190–197. doi: 10.1177/1468794112446106.

O'Rourke, M. and Crowley, S. J. (2013) 'Philosophical intervention and cross-disciplinary science: the story of the Toolbox Project', *Synthese*, 190(11), pp. 1937–1954. doi: 10.1007/s11229-012-0175-y.

Oliver, C. (2011) 'Critical realist grounded theory: a new approach for social work research', *British Journal of Social Work*, 42(2), pp. 371–387. doi: 10.1093/bjsw/bcr064.

Oman, S. (2019) 'Leisure pursuits: uncovering the "selective tradition" in culture and well-being evidence for policy', *Leisure Studies*, 39(1), pp. 11–25. doi: 10.1080/02614367.2019.1607536.

Oman, S. and Taylor, M. (2018) 'Subjective well-being in cultural advocacy: a politics of research between the market and the academy', *Journal of Cultural Economy*, 11(3), pp. 225–243. doi: 10.1080/17530350.2018.1435422.

Open Doors and Swarbrick, C. (2020) 'Editorial: Rebalancing the research agenda', *Dementia*, 19(1), pp. 3–5. doi: 10.1177/1471301219893134.

Palisson, J. *et al.* (2015) 'Music enhances verbal episodic memory in Alzheimer's disease', *Journal Of Clinical And Experimental Neuropsychology*, 37(5), pp. 503–517. doi: 10.1080/13803395.2015.1026802.

Parkinson, C., Windle, G. and Taylor, K. (2017) *Dementia & Imagination: research informed approaches to visual arts programmes*. Available at: http://www.artsforhealth.org/resources/dementia-and-imagination.pdf (Accessed: 14 January 2020).

Patton, M. Q. (1997) *Utilization-focused evaluation: The new century text*. 3rd edn. Thousand Oaks, CA: SAGE Publications Inc.

Pawson, R. (1996) 'Theorizing the interview', *The British Journal of Sociology*, 47(2), pp. 295–314.

Pawson, R. (2013) *The science of evaluation: a realist manifesto*. London: SAGE Publications Ltd.

Pawson, R. and Tilley, N. (1997) *Realistic evaluation*. e-Book. London: SAGE Publications Ltd.

Petticrew, M. *et al.* (2016) 'Å mêmes mots, sens différents - les difficultés de la terminologie épidémiologque avec la recherche en interventions en santé des populations', *Revue d'Épidémiologie et de Santé Publique*, (online 31.

Petticrew, M. and Roberts, H. (2003) 'Evidence, hierarchies, and typologies: horses for courses', *Journal of Epidemiology and Community Health*, 57(7), pp. 527–529.

Phillips, K. (2019) 'A constructive-critical response to The All-Party Arts, Health and Wellbeing Parliamentary Group report Creative Health: The Arts for Health and Wellbeing (July 2017)', *International Journal of Art Therapy*, 24(1), pp. 21–29. doi: 10.1080/17454832.2018.1491612.

Phillipson, L. and Hammond, A. (2018) 'More than talking: a scoping review of innovative approaches to qualitative research involving people with dementia', *International Journal of Qualitative Methods*, 17. doi: 10.1177/1609406918782784.

Pickett, J. *et al.* (2018) 'A roadmap to advance dementia research in prevention, diagnosis, intervention, and care by 2025', *International Journal of Geriatric Psychiatry*, 33(7), pp. 900–906. doi: 10.1002/gps.4868.

Pickett, J. and Brayne, C. (2019) 'The scale and profile of global dementia research funding', *The Lancet*, 394, pp. 1888–1889. doi: 10.1016/S0140-6736(19)32599-1.

Plsek, P. E. and Greenhalgh, T. (2001) 'Complexity science: the challenge of complexity in health care', *BMJ*, 323, pp. 625–8. doi: 10.1136/bmj.323.7313.625.

Porpora, D. V (1998) 'Four concepts of social structure', in Archer, M. (ed.) *Critical realism: essential readings*. e-Book. New York: Routledge.

Puurveen, G. *et al.* (2015) 'Ethical issues in the use of video observations with people with advanced dementia and their caregivers in nursing home environments', *Visual Methodologies*, 3(2), pp. 16–26. doi: 10.7331/vm.v3i2.49.

Ralph, N., Birks, M. and Chapman, Y. (2015) 'The methodological dynamism of grounded theory', *International Journal of Qualitative Methods*, 14(4), pp. 1–6. doi: 10.1177/1609406915611576.

Rapport, F., Wainwright, P. and Elwyn, G. (2005) '"Of the edgelands":

broadening the scope of qualitative methodology', *Medical Humanities*, 31(1), pp. 37–42. doi: 10.1136/jmh.2004.000190.

Raw, A. and Robson, M. (2017) 'Besieged by inappropriate criteria: arts organisations developing grounded evaluation approaches', in Reason, M. and Rowe, N. (eds) *Applied Practice: Evidence and Impact in Theatre, Music and Art*. London: Bloomsbury.

Richards, R. (2007) 'Everyday creativity and the arts', *World Futures*, 63(7), pp. 500–525. doi: 10.1080/02604020701572707.

Rose, G. (2014) 'On the relation between "Visual Research Methods" and contemporary visual culture', *The Sociological Review*, 62(1), pp. 24–46. doi: 10.1111/1467-954X.12109.

Sabat, S. (1994) 'Excess disability and malignant social psychology', *Journal of Community and Applied Psychology*, 4, pp. 157–166.

Sackett, D. L. *et al.* (1996) 'Evidence based medicine: what it is and what it isn't', *BMJ*, 312, pp. 71–72. doi: 10.1136/bmj.312.7023.71.

Sanderson, I. (2000) 'Evaluation in complex policy systems', *Evaluation*, 6(4), pp. 433–454. doi: 10.1177/13563890022209415.

Särkämö, T. *et al.* (2014) 'Cognitive, emotional, and social benefits of regular musical activities in early dementia: randomized controlled study', *The Gerontologist*, 54(4), pp. 634–650. doi: 10.1093/geront/gnt100.

Sayer, A. (1992) *Method in social science*. 2nd edn. London: Routledge.

Sayer, A. (2000) *Realism and social science*. London: SAGE Publications Ltd.

Schneider, J. (2018) 'The arts as a medium for care and self-care in dementia: arguments and evidence', *International Journal of Environmental Research and Public Health*, 15(6), p. 1151. doi: 10.3390/ijerph15061151.

Schön, D. A. (1983) *The reflective practitioner: how professionals think in action*. London: Ashgate Publishing Ltd.

Schwandt, T. A. (2007) 'Judging interpetations', *New Directions for Evaluation*, 114, pp. 63–81. doi: 10.1002/ev.

Scriven, M. (1967) 'The methodology of evaluation', in Tyler, R. ., Gagné, R. ., and Scriven, M. (eds) *Perspectives of curriculum evaluation (Vol. I)*. Chicago, IL: Rand McNally, pp. 39–83.

Selwood, S. (2002) 'The politics of data collection: gathering, analysing and using data about the subsidised cultural sector in England', *Cultural Trends*, 12(47), pp. 13–84. doi: 10.1080/09548960209390330.

Shakespeare, T., Zeilig, H. and Mittler, P. (2019) 'Rights in mind: thinking differently about dementia and disability', *Dementia*, 18(3), pp. 1075–1088. doi: 10.1177/1471301217701506.

Silvia, P. J. *et al.* (2014) 'Everyday creativity in daily life: an experience-sampling study of "little c" creativity', *Psychology of Aesthetics, Creativity, and the Arts*, 8(2), pp. 183–188. doi: 10.1037/a0035722.

Sixsmith, A. and Gibson, G. (2007) 'Music and the wellbeing of people with dementia', *Ageing and Society*, 27(01), pp. 127–145.

Skingley, A., Bungay, H. and Clift, S. (2011) 'Researching participatory arts, well-being and health: some methodological issues', *Journal of Arts & Communities*, 3(1), pp. 73–87. doi: 10.1386/jaac.3.1.73_1.

Skinner, M. W. *et al.* (2018) 'Improving social inclusion for people with dementia and carers through sharing dance: a qualitative sequential continuum of care pilot study protocol', *BMJ Open*, 8(11), pp. 1–7. doi: 10.1136/bmjopen-2018-026912.

Slay, J. and Ellis-Petersen, M. (2016) *The art of commissioning: how commissioners can release the potential of the arts and cultural sector*. London.

Spiro, N., Tsiris, G. and Cripps, C. (2018) '"Sounds good, but ... what is it?" an introduction to outcome measurement from a music therapy perspective', *Approaches: An Interdisciplinary Journal of Music Therapy*, (Online first 27 June 2018). Available at: http://researchonline.rcm.ac.uk/312/.

van Der Steen, J. *et al.* (2018) 'Music-based therapeutic interventions for people with dementia (Review)', *Cochrane Database of Systematic Reviews*, (7), pp. 1–86. doi: 10.1002/14651858.CD003477.pub3.

Stewart-Archer, L. A. *et al.* (2016) 'Subjective quality of life of those 65 years and older experiencing dementia', *Dementia*, 15(6), pp. 1716–1736. doi: 10.1177/1471301215576227.

Stickley, T. *et al.* (2017) ‘Arts, health and wellbeing: reflections on a national seminar series and building a UK research network’, *Arts & Health*. Taylor & Francis, 9(1), pp. 14–25. doi: 10.1080/17533015.2016.1166142.

Strauss, A. and Corbin, J. (1998) *Basics of qualitative research: techniques and procedures for developing grounded theory*. 2nd edn. Thousand Oaks, California: SAGE Publications Inc.

Sturges, J. E. and Hanrahan, K. J. (2004) ‘Comparing telephone and face-to-face qualitative interviewing: a research note’, *Qualitative Research*, 4(1), pp. 107–118.

Suri, H. (2013) ‘Epistemological pluralism in research synthesis methods’, *International Journal of Qualitative Studies in Education*, 26(7), pp. 889–911. doi: 10.1080/09518398.2012.691565.

Swaffer, K. (2014) ‘Dementia: stigma, language, and dementia-friendly’, *Dementia*, 13(6), pp. 709–716. doi: 10.1177/1471301214548143.

Swinnen, A. M. C. (2014) ‘Healing words: A study of poetry interventions in dementia care.’, *Dementia: The International Journal of Social Research and Practice*, 15(6), pp. 1377–1404.

Swinnen, A. M. C. (2016) ‘Healing words: a study of poetry interventions in dementia care’, *Dementia*, 15(6), pp. 1377–1404. doi: 10.1177/1471301214560378.

Swinnen, A. and de Medeiros, K. (2018) ‘“Play” and people living with dementia: a humanities-based inquiry of TimeSlips and the Alzheimer’s Poetry Project’, *Gerontologist*, 58(2), pp. 261–269. doi: 10.1093/geront/gnw196.

Taylor, T. (2013) ‘Well-being and prudential value’, *Philosophy and Public Policy Quarterly*, 31(2), pp. 10–17.

The Health Foundation (2013) *Quality improvement made simple: what everyone should know about health care quality improvement*. London: Health Foundation. Available at: https://www.health.org.uk/sites/default/files/QualityImprovementMadeSimple.pdf .

The Scottish Dementia Working Group Research Sub-group (2013) *Core* *principles for involving people with dementia in research - accessible version*. Available at: http://coreprinciplesdementia.files.wordpress.com/2014/05/dementia-a5-booklet.pdf.

Thomas, D. R. (2006) ‘A general inductive approach for analyzing qualitative evaluation data’, *American Journal of Evaluation*, 27(2), pp. 237–246. doi: 10.1177/1098214005283748.

Thomson, L. J. *et al.* (2018) ‘Effects of a museum-based social prescription intervention on quantitative measures of psychological wellbeing in older adults’, *Perspectives in Public Health*, 138(1), pp. 28–38. doi: 10.1177/1757913917737563.

Tobin, G. A. and Begley, C. M. (2004) ‘Methodological rigour within a qualitative framework’, *Journal of Advanced Nursing*, 48(4), pp. 388–396. doi: 10.1111/j.1365-2648.2004.03207.x.

Tronto, J. (1998) ‘An ethic of care’, *Generations*, 22(3), pp. 15–20.

UKRI (2019) *www.ukri.org.uk*, *Excellence with impact*. Available at: https://www.ukri.org/innovation/excellence-with-impact/ (Accessed: 20 September 2004).

University of Leeds (no date) *Centre for Cultural Value: Welcome*, *Centre for Cultural Value*. Available at: https://ahc.leeds.ac.uk/centre-cultural-value-1 (Accessed: 15 January 2019).

Urquhart, C. (2013) *Grounded theory for qualitative research: a practical guide*. London: SAGE Publications, Ltd.

Vella-Burrows, T. and Wilson, L. (2016) *Remember to Dance: evaluating the impact of dance activities for people in different stages of dementia*. Canterbury: Sidney de Haan Research Centre for Arts and Health.

Victor, C. *et al.* (2016) *Music, singing and wellbeing for adults living with dementia*. London: What Works Centre for Wellbeing.

Vink, A. C. *et al.* (2011) ‘Music therapy for people with dementia (Review)’, *Cochrane Database of Systematic Reviews*, (3), p. CD003477.

Walmsley, B. (2018) ‘Deep hanging out in the arts: an anthropological approach to capturing cultural value', *International Journal of Cultural Policy*, 24(2), pp. 272–291. doi: 10.1080/10286632.2016.1153081.

Wanzer, D. L. (2019) 'What is evaluation? Perspectives of how evaluation differs (or not) from research', *American Journal of Evaluation*, (Pre-print ahead of publication). doi: 10.17605/OSF.IO/H65ZM.

Webb, J. *et al.* (2020) 'Misfitting the research process: shaping qualitative research "in the field" to fit people living with dementia', *International Journal of Qualitative Methods*, 19. doi: 10.1177/1609406919895926.

Weber, M. (1949) *The methodology of the social sciences*. Edited by E. Shils and H. Finch. Glencoe, IL: The Free Press.

Weiss, C. H. (1979) 'The many meanings of research utilization', *Public Administration Review*, 39(5), pp. 426–431.

Weiss, C. H. (1998) 'Have we learned anything new about the use of evaluation?', *American Journal of Evaluation*, 19(1), pp. 21–33. doi: 10.1177/109821409801900103.

White, M. (2009) *Arts development in community health: a social tonic*. Oxford: Radcliffe Publishing.

Wieringa, S. *et al.* (2017) 'Has evidence-based medicine ever been modern? A Latour-inspired understanding of a changing EBM', *Journal of Evaluation in Clinical Practice*, 23(5), pp. 964–970. doi: 10.1111/jep.12752.

Wieringa, S., Dreesens, D., *et al.* (2018) 'Different knowledge, different styles of reasoning: a challenge for guideline development', *BMJ Evidence-Based Medicine*, p. ebmed-2017-110844. doi: 10.1136/bmjebm-2017-110844.

Wieringa, S., Engebretsen, E., *et al.* (2018) 'Rethinking bias and truth in evidence-based health care', *Journal of Evaluation in Clinical Practice*, 24(5), pp. 930–938. doi: 10.1111/jep.13010.

Williams, V. *et al.* (2020) 'Future lived experience: inclusive research with people living with dementia', *Qualitative Research*, (online first 6 January 2020), pp. 1–20. doi: 10.1177/1468794119893608.

Windle, G. *et al.* (2018) 'Exploring the theoretical foundations of visual art programmes for people living with dementia', *Dementia*, 17(6), pp. 702–727. doi:

10.1177/1471301217726613.

Windle, G. *et al.* (2019) 'Enhancing communication between dementia care staff and their residents: an arts-inspired intervention', *Aging & Mental Health*, (online first 18 March 2019). doi: 10.1080/13607863.2019.1590310.

Woolf, N. H. and Silver, C. (2018) *Qualitative analysis using NVivo: The Five-Level QDA® method*. New York: Routledge/Taylor & Francis Group.

Wuest, J. (1995) 'Feminist grounded theory: an exploration of the congruency and tensions between two traditions in knowledge discovery', *Qualitative Health Research*, 5(1), pp. 125–137.

Yeung, H. W. (1997) 'Critical realism and realist research in human geography: a method or a philosophy in search of a method?', *Progress in Human Geography*, 21(1), pp. 51–74. doi: 10.1191/030913297668207944.

Zeilig, H. (2016) 'Mark Making: methodologies and methods (innovative practice)', *Dementia*, 15(5), pp. 1295–1305. doi: 10.1177/1471301215595926.

Zeilig, H. *et al.* (2018) 'Co-creativity: possibilities for using the arts with people with a dementia', *Quality in Ageing & Older Adults*, 19(2), pp. 135–145. doi: 10.1108/QAOA-02-2018-0008.

Zeilig, H., Killick, J. and Fox, C. (2014) 'The participative arts for people living with a dementia: a critical review', *International Journal of Ageing and Later Life*, 9(1), pp. 7–34.

Zeisel, J. *et al.* (2016) 'Ecopsychosocial interventions in cognitive decline and dementia: a new terminology and a new paradigm', *American Journal of Alzheimer's Disease & Other Dementias*. doi: 10.1177/1533317516650806.

9 APPENDICES

9.1 Dissemination activities

9.1.1 Publications

Gray, K, Evans, S.E., Griffiths, A. & Schneider, J. (2018). ‘Critical reflections on methodological challenge in arts and dementia evaluation and research’, *Dementia*, 17 (6), pp. 775-784.

9.1.2 Conference presentations

Gray, K (2019). ‘Methodological challenges for evaluation’. UK Dementia Congress, 5 November, Doncaster

Gray, K., Evans, S., Bray, J., Garabedian, C. (2019). ‘Artists’ experience of working in care homes’. 48th Annual British Society of Gerontology Conference, 10-12 July, Liverpool.

Gray, K. & Lazenby, E. (2019). ‘“What is it about?”: illustrating research – a collaboration to narratively disseminate research findings. British Psychology Society Seminar Series: Telling stories in and about research, 8 July, Bristol.

Gray, K. (2018). ‘Arts and dementia: The challenges of evaluation’. Art, Memory & Aging: Building Dementia-Friendly Communities through the Arts, National Art Center, 8 October, Tokyo. (Invited speaker).

Gray, K. (2017). ‘Arts-based activities for people living with dementia: Are they particularly challenging to evaluate?’ 46th Annual British Society of Gerontology Conference, 5-7 July, Swansea.

Gray, K. (2017). ‘Thinking about complexity in arts and dementia evaluation’. Powerful Partners: Advancing Dementia Care through the Arts & Sciences, RSPH, 24 November, London. (Invited speaker).

Gray, K., Evans, S., Griffiths, A. & Schneider, J. (2017). ‘Arts-based activities for people living with dementia: Challenges for evaluation’. Seeds of Knowledge Postgraduate Research Conference, 29 June, Worcester.

Gray, K. (2017). ‘Methodological challenges in arts and dementia: A critical perspective on the literature’. 1st International Arts & Dementia Research Conference, RSPH, 9-10 March, London.

Gray, K. (2016). ‘Arts for people with dementia: Why is it so hard to show that it works?’, Arts and Health South-West Annual Conference, 24 November, Devon.

9.1.3 Seminars and workshops

Gray, K. (2020). ‘Evaluating arts and dementia’. Arts Uplift Suitcase Stories Symposium, 16 March, Worcester.

Cousins, E. & **Gray, *K.*** (2018) ‘Making meaning and capturing value across disciplines’, Arts and Dementia: Impact and Implementation, 20 September, Nottingham.

Gray, K. (2018). ‘Critical realist grounded theory?’, Institute of Health and Society Qualitative Research Methods seminar: Innovative Methodologies, 14 February, University of Worcester.

Gray, K. (2017). ‘Arts and dementia: Methodological challenge’. Soft Evaluations Workshop, 19 October, University of Wolverhampton.

9.1.4 Poster presentations

Gray, K. (2017). ‘Arts-based activities for people living with dementia: Are they particularly challenging to evaluate?’ Seeds of Knowledge Postgraduate Research Conference, 29 June, Worcester.

Gray, K. (2017). ‘How might an exploration of complexity help in the evaluation of arts-based activities for people with dementia?’ We need to talk about complexity, 13-14 June, Green Templeton College, University of Oxford.

Gray, K. & Lazenby, E. (2019). ‘Arts and dementia evaluation: addressing some of the challenges’. British Psychology Society Seminar Series: Telling stories in and about research, 8 July, Bristol.

9.2 Critical reflections on methodological challenge in arts and dementia evaluation and research

Submitted to Dementia: the international journal of social research and practice 14/07/17. Accepted 01/09/17.

Karen Gray
Association for Dementia Studies, University of Worcester, UK

Simon Chester Evans
Association for Dementia Studies, University of Worcester, UK

Amanda Griffiths
Division of Psychiatry and Applied Psychology, School of Medicine, University of Nottingham, UK

Justine Schneider
Institute of Mental Health, University of Nottingham, UK

Abstract

Methodological rigour, or its absence, is often a focus of concern for the emerging field of evaluation and research around arts and dementia. However, this paper suggests that critical attention should also be paid to the way in which individual perceptions, hidden assumptions and underlying social and political structures influence methodological work in the field. Such attention will be particularly important for addressing methodological challenges relating to contextual variability, ethics, value judgement, and signification identified through a literature review on this topic. Understanding how, where and when evaluators and researchers experience such challenges may help to identify fruitful approaches for future evaluation.

This paper is based upon a presentation on the subject given at the First International Research Conference on the Arts and Dementia: Theory, Methodology and Evidence on 9 March 2017.

Keywords

arts evaluation methodology methodological challenge epistemology methods complexity

Corresponding author

Karen Gray The Association for Dementia Studies, Institute of Health and Society, St Johns Campus, University of Worcester, Worcester WR2 6AJ, UK. Email: karen.gray@worc.ac.uk

Critical reflections on methodological challenge in arts and dementia evaluation and research

Two commonplaces about the arts and dementia: (1) it works, and; (2) this is hard to 'prove'. Arts organisations, museums and heritage bodies, and individual practitioners are delivering an increasing amount of arts activity designed to enhance the health and wellbeing of people living with dementia. There is a corresponding growth in research intended to strengthen the evidence base for this work and a requirement to tackle methodological challenges facing evaluators and researchers.

Methodology might be defined as the work that goes into identifying one's research position and methods. It has also been termed 'the description, the explanation and the justification of methods' (Kaplan, 1964, p. 18). Kaplan's definition is highlighted because it draws attention to the subjective and dialogical nature of methodological work. Choices about methods are made in social, political and historical contexts. It has further been suggested that research methods are essentially performative, that 'they have effects; they make differences; they enact realities; and they can help to bring into being what they also discover' (Law & Urry, 2004, p. 393) This paper reflects on some implications of acknowledging these aspects of methodological work through consideration of the challenges facing researchers and evaluators of arts-based activities for people with dementia. Key challenges were identified through a critical literature review for an ongoing project whose aim is to signpost solutions for evaluation research practice.

Despite promising and positive findings, many arts and dementia studies will be dismissed because of alleged lack of 'rigour' and poor methodological quality (Vink et al., 2011). Underpinning theoretical frameworks are said to be insufficient (Beard, 2011; Young, Camic, & Tischler, 2016). Causal mechanisms are not well understood (Windle et al., 2014). Persuasive economic evaluation appears elusive, perhaps because of a lack of research designs suitable for robust comparative cost analysis and able to satisfactorily address questions of causality (Fujiwara, Kudrna, & Dolan, 2014). Finally, evaluators and researchers have failed to access the direct experience of people with dementia (Zeilig, Killick, & Fox, 2014). Overall the concern is that funders, commissioners and policymakers remain unconvinced of the value of investment (Fraser, Bungay, & Munn-Giddings, 2014).

In facing such issues, and disagreement about the best approach to future evaluation and research practice, arts and dementia is aligned with the wider field of arts and health research where, despite some ambivalence about it from those who deliver arts activities, the dominant discourse remains that of the health sciences (Daykin et al., 2017; Fancourt & Joss, 2014). Greater methodological rigour is advocated by those who view the primary aim of research as investigating whether and how arts activity might be integrated within evidence-based practice. Robust experimental designs, including the randomised controlled trial (RCT), are seen as desirable (Moniz-Cook et al., 2011). Mixed method studies combine the strengths or compensate for the limitations of both quantitative and qualitative methods and are popular in the wider field of public health (Pluye & Hong, 2014). Standard qualitative methods are commonly used on their own in this field to explore nuanced implications of the effects of the arts. While it is sometimes thought that these methods will be less acceptable in health contexts, this may not reflect the views of all health commissioners (Goulding, 2014). Assessment of artistic quality, aesthetic experience, cultural contributions, process, social impact and economic value appear rare in the arts and dementia literature. Social or experiential effects of the arts on individuals with dementia have been explored using methodologies including ethnography (Swinnen, 2014), interpretive phenomenological analysis (Gregory, 2011), personal narrative analysis (Fels & Astell, 2011), and participatory critical arts-based enquiry (Dupuis et al., 2016). However, such distinctive approaches are seldom employed in evaluating the impact, outcomes or implementation of arts and dementia activities.

The challenges of conducting a literature review about methodological challenge

What methodological challenges are experienced by researchers and evaluators of arts-based activities delivered for and with people living with dementia? One difficulty with conducting research around this question is that challenges might not be visible except ‘in the gaps’ – through reflection on what cannot be said (perhaps because it isn’t considered publishable by a particular journal), or what has not been done in a study (maybe because it didn’t get past the ethics review process, or was too expensive). A reading of the limitations in arts and dementia research studies might lead us to believe that proper application of method is particularly problematic for positivist researchers. It is useful to remember that clear signposting of research challenges may not always be presented by those approaching the subject from other perspectives, where issues such as researcher ‘reflexivity’ will be addressed, but not necessarily framed as limitations.

The main inclusion criterion for the review was that literature should be a published, empirical study reporting on the impacts, outcomes or implementation of arts-based activity for people with dementia. The process of defining criteria became a microcosm for debates in the field in general. What is, and is not, ‘art’? What is ‘activity’? What constitutes ‘dementia’? It is worth noting that language matters greatly, and not just in the context of a literature review. A word used in one field or discipline may have different implications in another. The lack of a clear distinction between – for example – what constitutes ‘therapy’ or ‘activity’ in both arts and research practice has been identified as contributing to charged debate in arts and health, even placing an ‘unreasonable’ burden of expectation on arts practice (Broderick, 2011). Methodological terms, such as ‘intervention’, used without careful definition, may lead to mis-calibration of judgements about study quality and the evidence produced (Petticrew, Viehbeck, Cummins, & Lang, 2016).

Online database searches were conducted in February 2016. After application of exclusion criteria, hand-searching and consultation with others, 75 journal papers, referencing both quantitative and qualitative studies, were set aside for detailed reading in June 2016. Fourteen systematic or other reviews focusing on arts activity for people with dementia were added. A narrative approach for the review, informed by critical synthesis methods, was chosen as best suiting the topic (Dixon-Woods, 2006; Thomas & Harden, 2007).

Questions of methodological challenge

Because methodology is seen as a process of 'justification, description and explanation', key patterns of challenge identified through the review are presented here as questions, drawing attention to their dialogical nature.

1. *How do we account for complex contexts?*

Context has been defined as 'anything external to the intervention that may act as a barrier or facilitator to its implementation, or its effect' (Moore et al., 2015a, p. 2). Also in relation to implementation, Damschroder and colleagues (2009) apply the term to the broad set of interacting variables active in the circumstances surrounding an intervention, including its specific setting. Understanding the processes informing these interactions may help in assessment of fidelity and quality of implementation, clarification of causal mechanisms, and interpretation of the findings of evaluation (Moore et al., 2015). It has been suggested that attention to process more generally, particularly at the level of organisational barriers and opportunities for implementation of activities, would enable arts and health researchers to contribute more effectively to the development and sustainability of practice and research (Clift, 2012). As has been noted, this area appears under-developed in the emerging field of arts and dementia where projects are often studied at pilot or developmental stage. The reasons for this may be structural since, as Goulding (2014) has noted in relation to arts on prescription for older people, the delivery of arts projects is often shaped by the scarcity of long-term strategic funding and the prevalence of small-scale, local or opportunistic implementations.

Many types of contextual interactions of interest emerged from the review. These include: the setting where the arts activity is experienced, its physical space, routines and associated practices; time – including the timing of individual iterations of an activity and developmental stage of the intervention; the cultural, social and relational networks of participants; the personal and clinical histories, cultural demographies and aesthetic and other preferences of participants, and; the policies, practices and discourse surrounding dementia and dementia care. In the studies reviewed, these elements constantly interact with the artistic, technical and relational practices of individual arts practitioners and organisations.

Arts-based activities for people with dementia are often considered and evaluated as ‘complex interventions’. Advocates of a realist approach have suggested that while the influential Medical Research Council (MRC) guidance on the evaluation of complex interventions (Craig et al., 2008) discusses ways in which evaluation may be used to build theory and understand causal mechanisms, it makes insufficient mention of the role of context for implementation (Fletcher et al., 2016). It has separately been argued that commonly applied definitions of ‘complexity’ may not fully engage with understandings of the concept from a systems perspective Rickles, Hawe, & Shiell, 2007), and that such engagement could be useful in supporting effective implementation and evaluation (May, Johnson, & Finch, 2016; Pawson, 2013). A re-framing of the concept of complexity in the light of these insights might see evaluators treating an arts for dementia activity as a critical event occurring within a complex adaptive system (Hawe et al., 2009), or, as a complex intervention inserted into a complex system (Pawson, 2013). Such treatments would lead to a view of contextual variations as elements integral to successful implementation rather than problems for the experimental method.

The capture, documentation, analysis and reporting of descriptive detail about an arts activity is itself challenging. To this challenge, we could add that of dealing, methodologically, with uncertainty when attempting to understand causality within the complex systems in which arts for dementia activities are introduced.

2. *How do we address issues of value?*

Any choice of evaluation question is likely to be, as Weber suggests, rendered significant in light of ‘the cultural values with which we approach reality’ (Weber, 1949, p. 78). An evaluator of arts-based activities for people with dementia will make value judgements about the nature and quality of evidence required, arts and culture in general and the specific arts activity, and also about the condition of dementia, the people who experience it and dementia care practice. We may disagree about whether something has *intrinsic* value or whether the value is realised *instrumentally* through an action whose impact is felt by someone or something. Something may be ascribed a *quantity* of value, and we also engage in *acts of valuing*. An evaluator engages in one such act: judging how well some value, implicit in the theory informing a project, has been expressed through its implementation in practice. However, the reason why a particular set of values has been arrived at is rarely articulated in evaluation or research studies.

The differing focuses of the studies reviewed demonstrate that arts activities for people with dementia are variously valued because they:

- Provide an alternative to pharmacological and psychosocial interventions aimed at 'managing' the behavioural and psychological symptoms of dementia;
- Have specific and measurable cognitive, social or emotional benefits;
- Improve communication and interaction between carers and people with dementia or lessen the 'burden' of care;
- Enhance subjective wellbeing or quality of life of people living with dementia or their caregivers;
- Offer opportunities for personally enjoyable and meaningful activity;
- Challenge dominant narratives about dementia and constructions of the self in dementia;
- Enrich communities through cultural inclusion and opportunities for engaged citizenship.

While it has been shown that the arts may be a cost effective way to improve quality of life for older people (Coulton et al., 2015), the reviewed literature contained no economic studies.

The diversity and multiplicity of values identified, sometimes within a single study, suggests that a correspondingly rich and varied set of methodological tools is required to investigate them, and to unpick the implications of unreported value judgements or assumptions.

3. *What makes for an ethical evaluation?*

The involvement of persons living with dementia in research, including those lacking capacity, is an evolving issue. In the UK, the Mental Capacity Act 2005 provides researchers with a framework and code of practice (Department for Constitutional Affairs, 2007). The wishes of people with dementia are rightly recognised, alongside a requirement to support their continued individual participation in society. It has long been suggested that there are ethical questions to be answered around the delivery of arts activities effecting a change where a participating subject has not given, or is not able to give consent (Matarasso, 1996). If it fails to address such issues, evaluation research could be viewed as an activity ‘done to’ participants with dementia, with or without their consent (Bryden, 2015).

Thinking around personhood encourages us to view all behaviour of people living with dementia as meaningful. People with dementia might not lack capacity to communicate their opinions and feelings (Mozley et al., 1999) but there is little indication in the published literature that they are routinely consulted about or included in the development, implementation, or evaluation of activities. Participatory and wider stakeholder co-produced research processes may be of value for arts and health (Daykin et al., 2017) but appear uncommon within arts and dementia.

Institutional ethics processes are, moreover, described as cumbersome, time-consuming (a particular issue for evaluation of existing projects) and potentially exclusionary (Murray, 2013; Nuffield Council on Bioethics, 2009). Their processes, largely designed around clinical research models, may not sit well with low risk research and evaluation in arts and dementia contexts.

4. *How is meaning communicated and interpreted?*

Any exploration of our engagements in and with art will challenge us to think deeply about signification – the way in which meaning is conveyed and expressed. Art does not always give up its meaning easily, and the kind of meaning-making that takes place when we ‘do art’ happens at unexpected times and places: within the practice of the artist or in the production of an artwork, in a momentary interaction experienced by an individual engaging with art or with an arts practitioner or subsequently in the ‘story’ of that individual. Recognising, capturing and then making sense of significant moments so that they may be evaluated can be difficult, particularly when the person engaging with art has diminished cognitive ability, language skills and difficulties with attention (van Baalen et al., 2010). This point is illustrated in a project exploring art-viewing and making in a US art gallery in which a care partner reportedly told the researcher that a person with dementia had not understood what was going on during the art-viewing phase of an activity (Burnside et al., 2015). This might make us question what it means, methodologically, that someone engaging in an activity does not ‘understand’ that activity in the way that an evaluator (or an artist, or a care partner) does.

Researchers and evaluators report difficulties in using direct research methods with people with dementia (Cridland et al., 2016), some of which may originate in societal stigma. The effect of stigma in relation to participation in research has been documented (Batsch & Mittelman, 2012; Swaffer, 2014). One implication is that research activity might need to explore models of enquiry that embrace and empower communication and capability rather than cognitive failure or impairment.

Conclusions

Rigorous application of method is only one problematic area for research into arts-based activities for people with dementia. Challenges for documentation and analysis of the complex contexts in which activities are delivered also require consideration, not least because of how addressing them could inform the shape of future projects and our understanding of how change is effected. There may be an ethical imperative to find modes of enquiry which empower the communication capabilities of people living with dementia, and to explore ethical processes that more fully allow their informed involvement. In addition, evaluators and researchers may benefit from considering the impact for methodology of assumptions about value.

Stakeholder dialogue around the use of flexible methodological approaches may help us to understand better what happens when people with dementia engage with art, and to ensure that art is used effectively. Attempts to address the identified challenges could lead, for example, to the use of ethnographic, participatory or narrative approaches alongside implementation and process or realist evaluation methodologies. Questioning the assumed usefulness of hierarchies of evidence, Petticrew and Roberts remind us that “[e]nd point users, policy makers, and practitioners in particular ask many questions about interventions that are not just about effectiveness” (Petticrew & Roberts, 2003, p. 523). So, what other questions might we need to be asking, and who might we need to be listening to, in our drive to move on from the ‘it works, but it’s hard to prove it’ commonplaces?

These reflections on methodological challenge in arts and dementia suggest that further research could usefully explore how and why methodological choices are made, what pragmatic, political or personal factors determine these choices and the success of the resulting work, how other stakeholders are involved, if and how the cross-disciplinary nature of the arts and dementia field is reflected in the process, whether tensions result, and how these might be resolved. This may be particularly useful if, with Law and Urry (2004), we suspect that methods not only uncover realities, but also produce them.

Acknowledgements

The authors would like to thank those who offered responses and discussion at the RSPH First International Research Conference on the Arts and Dementia following the presentation upon which this paper is based. This work is supported by the Alzheimer’s Society, through funding of a doctoral studentship.

Conflict of interest

The authors declare that there is no conflict of interest.

References

Batsch, N. L., & Mittelman, M. S. (2012). World Alzheimer Report 2012: Overcoming the stigma of dementia. Alzheimer’s Disease International. London.

Beard, R. (2011). Art therapies and dementia care: A systematic review. *Dementia: The International Journal of Social Research and Practice*, *11*(5), 633–656.

Broderick, S. (2011). Arts practices in unreasonable doubt? Reflections on understandings of arts practices in healthcare contexts. *Arts & Health*, *3*(2), 95–109.

Bryden, C. (2015). Nothing about us without us! 20 years of dementia advocacy. London: Jessica Kingsley Publishers.

Clift, S. (2012). Creative arts as a public health resource: moving from practice-based research to evidence-based practice. *Perspectives in Public Health*, *132*(3), 120–127.

Coulton, S., Clift, S., Skingley, A., & Rodriguez, J. (2015). Effectiveness and cost-effectiveness of community singing on mental health-related quality of life of older people: Randomised controlled trial. *British Journal of Psychiatry*, *207*(3), 250–255.

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2008). Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ*, *337*, a1655.

Cridland, E. K., Phillipson, L., Brennan-Horley, C., & Swaffer, K. (2016). Reflections and recommendations for conducting in-depth interviews with people with dementia. *Qualitative Health Research*, *26*(13), 1774–1786.

Damschroder, L. J., Aron, D. C., Keith, R. E., Kirsh, S. R., Alexander, J. A., & Lowery, J. C. (2009). Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implementation Science*, *4*(1), 50.

Daykin, N., Gray, K., McCree, M., & Willis, J. (2017). Creative and credible evaluation for arts, health and well-being: opportunities and challenges of co-production. *Arts & Health: An International Journal for Research, Policy and Practice*, *9*(2), 123–138.

Department for Constitutional Affairs. (2007). *Mental Capacity Act 2005 Code of Practice*. London.

Dixon-Woods, M. (2006). How can systematic reviews incorporate qualitative research? A critical perspective. *Qualitative Research*, *6*(1), 27–44.

Dupuis, S. L., Kontos, P., Mitchell, G., Jonas-Simpson, C., & Gray, J. (2016). Re-claiming citizenship through the arts. *Dementia: The International Journal of Social Research and Practice*, *15*(3), 358–380.

Fancourt, D., & Joss, T. (2014). Aesop: A framework for developing and researching arts in health programmes. *Arts & Health: An International Journal for Research, Policy and Practice*, *7*(1), 1–13.

Fels, D. I., & Astell, A. J. (2011). Storytelling as a model of conversation for people with dementia and caregivers. *American Journal of Alzheimer's Disease & Other Dementias*, *26*(7), 535–541.

Fletcher, A., Jamal, F., Moore, G., Evans, R. E., Murphy, S., & Bonell, C. (2016). Realist complex intervention science: Applying realist principles across all phases of the Medical Research Council framework for developing and evaluating complex interventions. *Evaluation*, *22*(3), 286–303.

Fraser, A., Bungay, H., & Munn-Giddings, C. (2014). The value of the use of participatory arts activities in residential care settings to enhance the well-being and quality of life of older people: A rapid review of the literature. *Arts & Health: An International Journal for Research, Policy and Practice*, *6*(3), 266–278.

Fujiwara, D., Kudrna, L., & Dolan, P. (2014). Quantifying and valuing the wellbeing impacts of culture and sport. Department for Culture, Media & Sport Research Paper. London.

Goulding, A. (2014). Arts on prescription for older people: Different stakeholder perspectives on the challenges of providing evidence of impact on health outcomes. *Journal of Applied Arts & Health*, *5*(1), 83–107.

Gregory, H. (2011). Using poetry to improve the quality of life and care for people with dementia: A qualitative analysis of the Try to Remember programme. *Arts & Health: An International Journal for Research, Policy and Practice*, *3*(2), 160–172.

Kaplan, A. (1964). The conduct of inquiry; methodology for behavioral science. San Francisco: Chandler Publishing Company.

Law, J., & Urry, J. (2004). Enacting the social. *Economy and Society*, *33*(3), 390–410.

Matarasso, F. (1996). Defining values: Evaluating arts programmes. The Social Impact of the Arts. Stroud, Comedia.

May, C. R., Johnson, M., & Finch, T. (2016). Implementation, context and complexity. *Implementation Science*, *11*(1), 141.

Moniz-Cook, E., Vernooij-Dassen, M., Woods, B., Orrell, M., & INTERDEM Network. (2011). Psychosocial interventions in dementia care research: The INTERDEM manifesto. *Aging & Mental Health*, *15*(3), 283–290.

Moore, G. F., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., … Baird, J. (2015). Process evaluation of complex interventions: Medical Research Council guidance. *BMJ*, *350*, h1258.

Mozley, C., Huxley, P., Sutcliffe, C., Bagley, H., & Burns, A. (1999). 'Not knowing where I am doesn't mean I don't know what I like': Cognitive impairment and quality of life. *International Journal Of Geriatric Psychiatry*, *14*, 776–783.

Murray, A. (2013). The Mental Capacity Act and dementia research. *Nursing Older People*, *25*(3), 14–20.

Nuffield Council on Bioethics. (2009). *Dementia: Ethical issues*. London.

Pawson, R. (2013). The science of evaluation: A realist manifesto. London: Sage.

Petticrew, M., & Roberts, H. (2003). Evidence, hierarchies, and typologies: Horses for courses. *Journal of Epidemiology and Community Health*, *57*(7), 527–529.

Petticrew, M., Viehbeck, S., Cummins, S., & Lang, T. (2016). Å mêmes mots, sens différents - les difficultés de la terminologie épidémiologque avec la recherche en interventions en santé des populations. *Revue d'Épidémiologie et de Santé Publique*, published online 31 March 2016.

Pluye, P., & Hong, Q. N. (2014). Combining the power of stories and the power of numbers: Mixed methods research and mixed studies reviews. *Annual Review of Public Health*, *35*(1), 29–45.

Rickles, D., Hawe, P., & Shiell, A. (2007). A simple guide to chaos and complexity. *Journal of Epidemiology & Community Health*, *61*(11), 933–937.

Swaffer, K. (2014). Dementia: Stigma, language, and dementia-friendly. *Dementia: The International Journal of Social Research and Practice*, *13*(6), 709–716.

Swinnen, A. M. C. (2014). Healing words: A study of poetry interventions in dementia care. *Dementia: The International Journal of Social Research and Practice*, *15*(6), 1377–1404.

Thomas, J., & Harden, A. (2007). Methods for the thematic synthesis of qualitative research in systematic reviews (NCRM Working Paper Series No. 10). London.

van Baalen, A., Vingerhoets, A. J. J. M., Sixma, H. J., & de Lange, J. (2010). How to evaluate quality of care from the perspective of people with dementia: An overview of the literature. *Dementia: The International Journal of Social Research and Practice*, *10*(1), 112–137.

Vink, A. C., Birks, J. S., Bruinsma, M. S., & Scholten, R. J. (2011). Music therapy for people with dementia (Review). *Cochrane Database of Systematic Reviews*, (3), CD003477.

Weber, M. (1949). *The methodology of the social sciences*. (E. Shils & H. Finch, Eds.). Glencoe, Illinois: The Free Press.

Windle, G., Gregory, S., Newman, A., Goulding, A., O'Brien, D., & Parkinson, C. (2014). Understanding the impact of visual arts interventions for people living with dementia: A realist review protocol. *Systematic Reviews*, *3*(91).

Young, R., Camic, P. M., & Tischler, V. (2016). The impact of community-based arts and health interventions on cognition in people with dementia: a systematic literature review. *Aging & Mental Health*, *20*(4), 337–351.

Zeilig, H., Killick, J., & Fox, C. (2014). The participative arts for people living with a dementia: a critical review. *International Journal of Ageing and Later Life*, *9*(1), 7–34

9.3 Literature review: Table of characteristics of published studies

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
Included studies from searches in 2016				
[1] (Argyle and Kelly, 2015) UK.	To examine potential benefits and barriers and facilitators to implementation of personalised musical intervention delivered by home care workers during visits to clients with dementia.	Qualitative interviews. Thematic analysis.	Intervention not implemented as intended.	Identifies gap between research evidence and practical implementation of person-centred interventions as key issue. Discusses knowledge translation and problems of bridging gap between evidence and practice.
[2] (Bannan & Montgomery-Smith, 2008) UK.	To report on a pilot study of 3 weekly sessions of group singing for people with Alzheimer's Disease and their carers.	Mixed. Video and audio recordings of sessions. Questionnaires completed by carers and health workers.	Little direct questioning of participants possible. Reliance on opinions of carers and health workers. 'Desire to please' in carers and health workers.	
[3] (Billington et al., 2013) UK.	To investigate the effects of a shared reading intervention on BPSD Looked at different environments, identifying staff perceptions of influence.	Mixed method service evaluation study. 3 healthcare environments (care homes, hospital wards, day centre). NPI-Q for staff views of changes in symptom severity in care homes. Interviews with staff (not recorded or transcribed). Thematic analysis. n=61 PWD, n=20 staff. Member validation with staff of thematic analysis.	Involved three different environments, and different care homes and wards. Data collected at different times: one week before baseline, then one care home waited three times (termed 'a wait list design'), others held intervention groups continually over 6 months. Day centre continued for 8 months, hospitals over 12 weeks. Different stages of dementia. Some being assessed or recently diagnosed. Many comorbidities. Staff excluded participants they felt not appropriate, or if individual didn't want to attend.	Recommends RCT with qualitative component. Suggests longitudinal observational/case study design. Suggests comparing different methods of reading intervention.

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
			Limited sample size (reasons for non-completion included death, moved from home, entered care home, left hospital, no baseline). Impact of individual differences –no knowledge of prior experience, medication use, or participation in other interventions.	
[4] (Burnside et al., 2017) USA.	To explore the benefits of an arts engagement program in an art gallery for PWD and care partners, and to use results to develop conceptual model for assessing outcomes.	Qualitative evaluation and grounded theory. N=21 (care partners), n=13 PWD interviewed (telephone). Demographics and consent prior to first session.	Self-selected convenience sample, no randomisation. Homogenous, educated, mostly white, with prior interest / experience. Most mild/early stage dementia. No medical chart reviews. Type and severity of dementia not clinically assessed. Care partners responsible for rating cognitive and functional ability using CDR scale. Telephone interviews difficult for PWD with verbal communication problems. No direct comparison between tours and classes, so couldn't compare dose response or specific effect of artmaking Participants interviewed within 2 weeks rather than immediately after for logistical reasons, not to cause fatigue, to maintain separation between experience of program and study.	Identifies gap in understanding specific benefits and components of programs in cultural arts venues. Authors suggest qual methods bring 'voice and perspective' of PWD, adding to quant findings of other studies. Suggests 'process not be underestimated' in favour of outcomes. Recommends future studies assess dose, investigate how to accommodate progressive impairment, and explore effects of different media.
[5] (Byrne & Mackinlay, 2012) Australia.	To explore the effects of a weekly person-centred 1-hour art-making workshop on participants with dementia and depression, over an 18 month period.	Qualitative component of study of n=15 PWD who were also depressed. Sessions audio-recorded and transcribed. RAs kept journal of non-verbal behaviours. Facilitator reflection journal kept. NViVo.	Participants found completing smiley face scales difficult and time-consuming. Small sample size. Care staff didn't fully understand the intervention. Expectations may have guided interactions and responses. Values and assumptions about art and people involved may have affected communication of project by facilitators.	Data analysis failed to show benefits, staff believed residents did benefit. Authors note need for variety in sources. Suggest video recording for future, but note expense.

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
[6] (Camic, Williams, & Meeten, 2011) UK.	To evaluate whether a 10-week singing together group intervention improves QoL of PWD dementia and carers.	Pilot evaluation study using mixed methods and repeated measures. Mood, QoL, BPSD, activities of daily living and cognitive status measured pre, post and 10 weeks after. Engagement levels monitored during each session, care partners rated each session. Interviews (dyads separated) pre-post – thematic analysis. n=10 PWD and carer dyads. Degree of dementia assessed with ACE-R at start and MMSE at follow-up. NPI. GDS, DEM-QoL-4 and proxy version. BADLS. Carers DAS and QoL (WHO-QoL BREF). Specially constructed observational scale during session. Carers kept weekly diary of singing and music-listening throughout.	Only 7 of 10 dyads available for 10 week follow-up (reasons included health deterioration and recovering from accident). Only pre-post measures further analysed. PWD fatigue or difficulties with some scales. Results from standardised measures inconclusive, may reflect ongoing difficulty in adequately measuring QoL in deteriorating condition. Difficulty in obtaining reliable standardized scores on QoL / other psychosocial constructs in PWD with moderate/severe impairment. Participants had high baseline scores. Results showed no significant differences in PWD QoL, but depression increased. Gradual deterioration in cognitive abilities, activities of daily living and BPSD. A change in medication affected one dyad. No significant difference in carer QoL seen, although this may be because study included people who were already coping well. Small scale study, short time frame. Needed comparison.	Phase 1 exploratory study (cf MRC guidelines). Worked on basis that QoL was associated with BPSD – greater BPSD, lower QoL. Notes lack of attention to carer's sense of self – wellbeing may not be associated with PWD wellbeing. Recommend future studies should be larger scale (preferably RCT), include health economic analysis, and examine wide range of activities. Need to explore observation that reduction in negative mood or experiences may not result in increase in wellbeing or happiness.
[7] (Camic, Tischler, & Pearman, 2014) UK.	To understand experience of 8-week art-gallery-based intervention at 2 different galleries for people with mild/moderate dementia and their carers. To examine impact on social inclusion, carer burden, QoL and daily living activities for PWD.	Mixed methods pre-post, using standardised questionnaires and interviews. Compares similar interventions at traditional and contemporary art galleries. n=24, (n=12 PWD). Dem-QoL 4. ZBI and BADLS. Thematic analysis. Interviews in participants' home. Also detailed researcher field notes.	Standardised measures problematic with small sample size. Detection of no change could be due to: lack of measure specificity in relation to type of intervention; expressed preferences of participants to talk about experiences with a person rather than respond to questionnaire; small sample size. No control group. Results showed no difference between gallery sites, but without control group, difficult to fully interpret. Although qual results lend support to analysis, can only be assessed in the particular study context.	No significant pre-post difference between two groups on quant measures. Non-significant trend towards reduction carer burden in both. Wellbeing benefits emerged from qual. Recommends audio/video recording of sessions, with quant and qual analysis to understand subtle changes.

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
			DEMQoL-4 showed no significant change, self-and carer reports indicated otherwise. Themes in qual analysis were only those relevant to all participants, information may have been omitted that influenced author's interpretation – although negative or critical comments not omitted.	Suggests RCT with active control and usual care groups, but notes standardising intervention over multiple institutions likely impossible. Recommends natural experiment design in different contexts, making use of existing programmes.
[8] (Camic, Baker, & Tischler, 2016) UK.	To explore the experience of PWD and caregivers taking part in 8 2-hour group sessions (1 hour art viewing, 1 hour art-making) in 2 art gallery sites over an 8 week period, and to develop a theory to allow better understanding how programs at art galleries might play role in lives of PWD.	Qualitative. Grounded theory. n=12 PWD, n=12 caregivers, n=4 gallery facilitators. Interviews, field notes and communication between facilitators and researchers. ACE-R, NPI-Q for inclusion. Detailed observational field notes.	Sample only included PW mild/moderate dementia over age of 55. Participants were self-selected, may have prior interest. Facilitators able to clarify gaps or discuss ideas throughout, participants could not – potentially limiting aspects of theory development.	
[9] (Chen et al., 2016) China.	To assess the effect of TimeSlips creative expression intervention on depression for PWD with mild to moderate dementia.	Quantitative. Pre-post (one week and one month after). CSDD and OERS. n=43. Drop-out rate 12% from initial n=49 recruited. MMSE (Chinese version) used to assess cognitive function.	Need comparator to be able to say whether intervention or simply non-drug therapy caused effect. Subjective scoring – caregivers might score more highly because they see any improvement as highly positive.	Found significant positive differences on both scales used. Recommends large scale experimental study to follow up. Out of 49 initial patients, 6 withdrew during process, and 3 for disease progression. Another 3 dropped out.
[10] (Clement et al., 2012) France.	To assess the short and longer term effects of 4 week, twice-weekly, musical and cooking interventions on the emotional	Randomised trial, repeated measures, pre and post. Two matched groups of PWAD (n=14). Repeat evaluation of emotional state before, during	Attrition and small sample size – 3 of 14 didn't complete (poor health and death). Not possible to say whether benefits are specific to music or relate to auditory stimulation in general.	Suggests exploration of positive short term effect of music on emotional state and long term positive effect of music on mood regulation.

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
	wellbeing of people with severe Alzheimer's disease.	and after. Discourse content and facial expressions from short filmed interviews, along with caregiver's judgements of mood. Baseline measures 1 week before and on first day. Post one hour and day after last session. Then at + 2 and +4 weeks. Interviews carried out by caregiver and filmed.		Musical experience questionnaire showed experience was low and none had specific cooking experience. No discrepancies in cognitive, behavioural or demographic data between 2 groups.
[11] (Cooke et al., 2010a) Australia.	To assess the effect of participation in a 40 min live and interactive group music programme 3 times per week for 8 weeks, on agitation and anxiety in older PWD, compared with a reading group.	Randomised cross-over design with music intervention and reading control group. n=47 mild / moderate dementia in 2 aged care facilities. AMAI-SF, rated by care staff blinded to treatment group. RAID (anxiety), self-rated, administered by RAs blinded to treatment groups. RAs assessed dementia severity using MMSE and collected demographics at baseline, along with musical preferences (using an adapted Music Preference Questionnaire). Engagement observation checklist devised by research team.	Notes lack of agreement on how to define construct of anxiety. Participants identified by care facility managers, based on inclusion criteria and randomised by blinded biostatistician. Participants given option to attend or not before each session, leading to differences in attendance levels. Participants may have been doing similar activities concurrently. Notes participants' consistently low levels of anxiety and minimal displays of agitated behavior. Notes possibility that music may not have greater therapeutic effect than reading, and that both groups offered greater engagement than routine care, perhaps leading to the lack of significant difference. Usual care control group would have helped to contextualise. Notes possibility that interventions might need to be individualized and that some participants might have responded better to music, some to reading. Outcome measures completed by different kinds of respondents. Agitation rated by carers, anxiety assessed through self report. Future studies should use the same kind of respondent.	Did not significantly affect agitation and anxiety. Both activities did give some participants a 'voice', increasing verbalization. Analysis followed intention-to-treat principle. Noting low baseline scores, authors recommend that study inclusion should be based on measure to be used in study, not documentation or nursing records Music may have only short term effects, future research may need to be conducted when symptoms are most prevalent.

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
[12] (Cooke et al, 2010b) Australia.	To assess the effect of participation in 40 min live and interactive group music programme 3 times per week for 8 weeks, on QoL and depression in PWD, compared with a reading group.	RCT cross-over design with comparison reading group. N=47 PWD, all able to verbalise answers, from two care facilities in Australia.. Randomisation by computer by blinded biostatistician. Wash-out period included to reduce potential carryover effects. Measures included DQoL and GDS at baseline, mid-point and post. RAs blinded to treatment groups.	Considerable efforts made to ensure treatment fidelity. Consistently low levels of depression and good QoL at baseline and throughout. Low baseline scores. Maybe effects of music are short term and dissipate immediately after. Differences in delivery of sessions. Music very structured, much more facilitator-led. Reading sessions more organic with opportunities for greater individual involvement. Would explain feelings of greater 'belonging'.	Treatment fidelity: standardised procedures manual developed and facilitators trained in delivery of sessions and in working with PWD, practice session and random spot checks. Musical preferences and experiences assessed. Consider other screening methods to assess baseline levels prior to study. Usual care group would be useful.
[13] (Davidson & Almeida, 2014) Australia.	To examine the effect of group singing activity on the lucidity, mood, agitated behavior and focus of PWD, and caregivers' energy, mood, level of stress/relaxation and focus.	Mixed methods. Exploratory study in 2 stages. Stage 1: one-off structured singing session compared with verbal quiz control activity. Participants had mild/ moderate dementia and were from residential care and community contexts. Quant measures for PWD and caregivers. Stage 2: 6 week musical engagement intervention with community based participants. Quant data, 2 hours before and 2 hours after. Supplemented with informal semi-structured interviews to engage participants in stage 2.	Difficulty in developing suitable control activity containing same kinds of elements as intervention. Study authors wanted to preserve ecological validity – replicating norm of delivering session just once a week as usually happens outside research. Small convenience sample. Reasons for small size, including one group scheduled immediately after lunch, making it difficult to access one prof carer per PWD; drop-in centre was very small, couldn't recruit more than 12; varied attendance across programme. Rating scales developed specifically for study. Quant data provided only by proxy, with professional caregivers often reporting on several PWD. Professional carers have limited time to contribute. Time demands on participants meant only possible to collect data at weeks 2, 4 and 6.	Authors had previously developed the structured singing activity. Same facilitator delivered both singing and quiz activities in stage 1. Brief qual interviews said to offer 'richer insight' than quant results. Qual data supported and supplemented, offering richer and 'more human face' to rating scale results.

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
		Facilitator contributed reflections.	Participants (stage 1) from different contexts – drop in day centre and residential care. Data analysed for each group and combined (meaning that contextual and other confounding variables lost). Because stage 1 didn't show signs of significant improvement across all variables in the quiz group, authors didn't run control group for stage 2.	
[14] (Davidson & Fedele, 2011) Australia.	To assess the effects on mood and quality of life of a 6 week singing program for older PWD and caregivers at residential care unit and drop-in program in community.	Mixed methods pre / post. QoL measures, self and proxy (QoL-AD) Measures of mood, energy, concentration and stress level for caregivers. HDS. Observations and checklist. Two settings: community (n=23, 11 PWD, 11 caregivers, 1 caregiver on own); care home (n=25, 18 PWD, 7 support workers). Video-taped sessions.	Results showed no significant differences on QoL or between pre and post HDS scores. Qual observations more positive, demonstrating engagement during sessions. Suggests insufficient time for impact to be quantitatively beneficial or for decline in dementia to occur. Notes difference between responses from caregivers and PWD – caregivers thought living situation of PWD had improved. Suggests participants may have been responding to perceived demand of study in order to appear 'normal' or to please and that measures may have lacked specificity.	Recoded negatively worded scale items on all measures. Results showed immediate energizing effect, but no longer term impact. Living situation of PWD did significantly improve – in views of caregivers. Notes previous studies have not considered group history, regular program of activities undertaken or what types of people attend – making it difficult to generalise. Facilitator had not worked with dementia-specific singing group previously.
[15] (van Dijk, Van-Weert, 7 Dröes, 2012) Netherlands.	To assess the use of a living room theatre communication activity for people with dementia in psychogeriatric nursing home settings.	Quasi-experimental 3 group design. Group 1, n=65 PWD experienced living room theatre activity given by caregivers; Group 2, n=31, activity by actors; control, n=55, usual reminiscence group activity. Participants drawn from 22 wards across 13 nursing homes. Scales:	Group 2 longer duration intervention than Group 2. Caregivers delivering intervention reported feeling nervous due to research observations and presence of cameras. Interruptions during caregiver led group – eg dentist arriving, building works. Authors noted differences between nursing home wards and between residents that led them to globally match nursing homes. Also differences between groups on baseline characteristics,	Efforts made to match participant groups in each group, based on MMSE, composition of nursing staff and care model used in wards. Control groups had to take place in settings where no Veder method staff had been trained.

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
		behaviour (adapted INTERACT), mood (FACE expression) and QoL (adapted QUALIDEM), involvement with environment (4 point Likert scale derived from Dutch DCM) measured using standardised observation scales pre, during and just after intervention. Single perceived QoL question verbally pre and post-test. Application of Veder method measured using specially developed instrument.	included in analyses as covariates. Variety in severity of cognitive impairment. Post-test measurement performed shortly after intervention and no long term effects studied. Low numbers of participants, and not equally distributed, partly due to lack of time to organize sessions. Under-powered, generalisability affected. Inter-rater reliability of some items and subscales low. The second observations were not performed live, but by watching video recordings afterwards.	Sample size determined by power calculation. Needed 64 people for each group, but didn't achieve. Observers not informed about goal of research and had no prior information about which condition they were observing.
[16] (van Haeften-Van Dijk, van Weert, & Droës, 2015) Netherlands.	To evaluate the implementation of care staff training to apply the Veder living room theatre communication method on psychogeriatric nursing home wards. To explore the barriers and facilitators in different stages of the implementation process.	Process evaluation. Interviews with purposively sampled stakeholders (n=12) and focus groups with care staff (n=35 from 21 different nursing homes). Implementation Process Evaluation framework used to categorise data and 7s-model to contextualize qual findings. Analysis deductive, predetermined analysis scheme derived from IPE framework.	IPE Framework is complex – integration of two theoretical models in one. Selection bias – 4/5 focus groups conducted during 'refresher days', only one organized with people trained in the method but no longer applying it. Enthusiasts over-represented. Evaluation focused mainly on staff training as issues around sustainability were the responsibility of the nursing home organisation.	Comments on scarcity of studies on how to implement interventions.
[17] (Dupuis et al., 2016) Canada.	To explore issues of citizenship and how PWD and family members wished to be represented to the world.	Interdisciplinary, critical arts-based enquiry approach. 8 PWD, 15 artists, 7 researchers + volunteers. One day workshop. Audio and	None reported.	Participants and researchers talked about potential for project to expose vulnerabilities, but this is seen in a positive light.

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
		video-recorded focus groups followed by PWD working with artists to co-create artistic reflections. Post workshop telephone interviews analysed using critical creative hermeneutic analysis.		
18 (Eekelaar, Camic, & Springham, 2012) UK.	To evaluate the effects on cognition, specifically episodic memory and verbal fluency, for PWD taking part in a programme of structured viewing and discussion of paintings in a gallery, followed by art-making led by art therapist.	Exploratory study. Mixed methods pre-post with 4 week follow up. Participants (n=6 with their family carers), mild to moderate dementia (assessed using MMSE). Baseline established at home visit before and at 4 week follow-up – assessed through discussion of high quality art reproductions, audio-recorded). Audio recordings of sessions analysed for verbal fluency and episodic memory (quantitative content analysis). Family carers interviewed post intervention (thematic analysis). Open-ended semi-structured interviews.	Did not aim to determine a causal link between intervention and cognitive functioning and therefore conclusions as to association only speculative. Components other than art may have affected changes (eg social contact). Control group would have been helpful in establishing causal links, perhaps through comparison with similar activities in ordinary setting. Some dyads didn't attend all sessions. No type of dementia specified – different subtypes might respond differently.	Authors recognise importance of understanding underlying psychological mechanisms related to why PWD have responded positively to gallery-based interventions in order to further develop and refine. Method chosen for its qual and quant capabilities. Inter-rater reliability established. Did not use standardised psychometric tests because it was felt they would not be likely to detect small changes over short time-frame. Would also have been difficult to administer, have varying reliability and validity, and lengthy questionnaires don't work well for those with cognitive difficulties.
19 (Flatt et al., 2015)	To describe the subjective experiences of older adults with early stage AD and caregivers	Mixed methods. Focus groups following activity. Self-report satisfaction	Small self-selecting sample, limited to those with own transport to museum and physically able to participate in guided tour. Many were Alzheimer's	Reported on practical barriers and facilitators for future activities.

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
USA.	attending a one-time art museum engagement activity (discussion and art-making) at a museum.	survey Likert (non-parametric statistics). n=20 participants (n=10 PWD, n=10 carers). Field notes at focus group, and interviews thematically coded.	Disease Research Centre members. Majority had past experience visiting art museums or creating art. Notes possibility of ceiling effect in results from satisfaction survey.	Suggests more objective measures needed and notes few assessment tools available to evaluate impact of art-related interventions
[20] (Fletcher & Eckberg, 2014) USA.	To evaluate the effects of occupational therapist guided creative reminiscence activities on quality of life of 'clients with memory loss' attending a private day centre, and on caregivers' burden.	Pilot study. Quasi-experimental repeated measures. Standardised measures. 4 weekly hour long control activities, 4 week break, 4 week creative reminiscence. Observation to assess wellbeing using GCCW-BOT in both. QOL-AD. Caregivers reported on perceived burden (Zarit Burden Interview) and rated own QoL. Focus group 6 months after for caregivers in which findings were presented and discussed. Thematic analysis	Participants required significant time to complete measures, e.g. 15-20 minutes to read and respond to QoL-AD, longer than 10 minutes suggested. Most comfortable using helper to complete. ZBI took 15 minutes. Participants reported process tiring. Small sample size, so nonparametric statistics. Attrition from n=36 interested, to only n=12 completing intervention. Reasons: progress of condition, distracting other participants, death, inability to provide photographs for intervention. Intention to treat used to predict missing scores. Sample homogenous. Results showed no clear patterns of change in self-reported wellbeing or QoL, but caregiver perceived QoL increased over course of reminiscence activity with reduced feelings of burden. Many subtle events not recorded or measured by quant instruments. Participants questioned construct validity of questionnaires, saying information requested varied day to day, moment to moment. They suggested researchers and professionals 'don't know what it's like' and that QoL is perceived differently by caregiver and client. Participants emphasized importance of ongoing communication between researchers and participants.	No clear patterns of change seen in wellbeing overall or in subdomains. No clear patterns of change seen in QoL. Authors question whether QoL measures used can measure changes occurring over time and change as a result of treatment. It may be that people with dementia just reply in an unvarying way and may not refer to events or changes in the present. Behaviours of PWD themselves may themselves be unvarying, and restricted range of observed circumstances might result in restricted observation of wellbeing. Also questioned whether intervention was too structured for the GCCWOT to work well.

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
			Longer time period for intervention may have provided clearer data. Follow up didn't record longer term benefits to PWD. Intervention featured only one setting with one kind of activity and facilitator.	
[21] (Fritsch et al., 2009) USA	To evaluate the impact of TimeSlips creative expression programme on PWD in a nursing home. To assess the impact on care staff.	Observational study using experimental design. 20 nursing home facilities, matched pairs of 10 randomly selected to implement 10-week intervention and 10 control homes. Workshop and training of staff implementing intervention. 2 weeks after implementation, conducted 4 days of direct obs of residents and staff, using time sampling. Also surveyed staff job satisfaction, attitudes towards residents and burnout. Compared to non-TS facilities.	Nursing home facilities were convenience sample of those that volunteered to participate. Matched to control /intervention by number of beds, percentage Medicaid residents and location. Lack of impact on care staff attitudes, job satisfaction etc may be because intervention was short term, probably not long enough to really effect changes in job satisfaction/burnout etc. Also, localized and too small to overcome other factors affecting job satisfaction. Limited resources meant authors used time sampling but couldn't identify in advance which residents would provide data or limit observations to only residents exposed to full programme. Post only study design – no baseline. TS may have caused residents to be more active – potentially causing more 'challenging' behaviours.	Authors developed adapted study specific surveys and engagement observation measure. Training program is highly structured and manualized, helping to ensure consistent implementation. Recommends RCT with residents as unit of analysis, rather than nursing home.
[22] (George & Houser, 2014) USA.	To explore the experience of PWD residents in a nursing home taking part in a 6 week TimeSlips creative expression programme.	Qualitative. 3-5 minute interviews with (n=10) residents of nursing care unit taking part and 15-20 min interviews with staff. Thematic analysis. Part of cluster-randomised pilot study comparing 2 dem care units in nursing home (see above).	Intervention cohort all female with racial demographic imbalance. Interviews in one facility. No demographic or other data / characteristics collected from staff. During interviews, two residents silent or non-verbal. Interviews very short (didn't want to burden). Information may have been more robust if elicited throughout intervention. Could have included opportunity for residents to discuss things they hadn't enjoyed.	Recommends control using a conventional therapy in future studies. Would have been useful to include observational data on non-verbal communications of participants during activity. Need larger trials of longer duration to confirm qual findings.

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
			Demand characteristics may have played a role, particularly in relation to interviews with staff. All data pooled in single document.	
[23] (Gregory, 2011) UK.	To explore the impact of a poetry intervention on the quality of life and care of PWD at one care home and one day centre on PWD QoL and care. To evaluate the success of the project in meeting its aims.	Qualitative. IPA. Care staff (n=6) interviewed about experiences of project.		Notes potential for overestimating impacts. Discusses the limitations to successful implementation of projects. Suggests challenge now is to ensure findings are heard, to close gap between theory and practice. Kitwood argument that person-centred interventions cannot be evaluated the same way as drug treatments.
[24] (Gross McAdam, 2013) Australia.	To explore the relationship between art, dementia and wellbeing for PWD as part of an art-based community research project in aged care facilities.	Phenomenological study, exploring lived experience of n=12 residents taking part in creation of large-scale collaborative artwork. CEAA tool. Developed MAC.ART wellbeing assessment tool – matrix used to document emotional responses of participants to activity. Photographed and videoed and re-scored against CEEA.	None specifically noted.	Measuring 'here and now' activities, not using biomedical research outcomes. Participants are very different – lots of types or combinations of dementia, other disabilities, influence of medication, varied cognitive impairment levels. Recommends development of culturally appropriate assessment tools.
[25] (Gross et al., 2015) USA.	To evaluate whether there were carry-over effects for people with dementia resident in a care setting taking part in a 12-week	Mixed methods. Observation of 1, 6 and 12th session using GCCWOT for n=76 PWD in four long-term care facilities by interns. Trained care	Missing data. No control. Self reports unavailable. Ratings cannot be taken at face value. Perhaps caused by habituation of interns to behaviours of	Some criticisms of GCCWOT tool: Indications that it actually measures only Well-Being and Ill-Being? Rating scales could be clearer.

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
	art activity programme, and for the staff supporting them.	facility staff members assessed once on or close to same days outside art session. All participants late-middle to advanced stages. Intern journal entries.	residents over time? Staff already habituated. Using different observers inside and outside sessions. Art activity program designed for higher functioning individuals than those participating. Journal entries insufficient for rigorous qual analysis but provide anecdotal support.	More training for observers needed. Recommends for future: highly trained observers familiar and comfortable with dementia; evaluate one study participant at a time; use more than one instrument; complement with video samples; measurements more frequently during / following activity
[26] (Guzmán-García, Mukaetova-Ladinska, & James, 2013) UK.	To investigate the effect of introducing a dance-based psychomotor intervention using Latin ballroom for PWD in care homes. To use findings to design quantitative study.	Grounded theory qualitative pilot study. Interviews with n=7 PWD, n=9 staff and facilitators.	Small number of participants interviewed (although saturation reached). Not all dancers with dementia able or willing to participate in interviews. First author had dual role as researcher-therapist – facilitating the intervention. Qual study allows exploration of themes prior to further quant study and analysis.	Participants recruited on basis of recommendations by care home managers. Assessed by consultant old age psychiatrist, not taking part in any other activity programmes. Qual methodology allows participants' opinions to be recorded.
[27] (Hara, 2011) UK.	To explore individual experience of people with dementia part in a singing / music activity (Singing for the Brain) and how music might be used to support everyday lives.	Mixed methods. Extended ethnographic participant observation, interviews with participants. 12 sessions observed over 6 months, followed by further contact and participation in sessions. Used active music making as elicitation tool during interviews.	Discussed difficulties of approaching those with dementia to obtain data, particularly in advanced stages. Notes practical difficulties taking notes during 'covert observation' – tried not to show 'researcher' pose or to interrupt or influence sessions. Reflects on difficulties making sense of words of informants with dementia, notes author familiarity with individuals allowed more detail. Notes need to be conscious of normative expectations researcher bring to encounter – e.g. that you can't just passively 'listen' to PWD, you	Suggests medical and neuroscience model leaves sociocultural context of music listening out of analysis and notes that the situational experience of music by people with dementia is difficult to generalise in causal terms. In depth understanding of local context and connections experienced by participants. Didn't want to make interviews a burden for participants,

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
			have to relate to and communicate proactively in order to elicit responses. Music used to do this.	careful to consult with carers as to appropriate time and context Notes importance of setting for interview being somewhere both carer and participant feel comfortable and able to be spontaneous.
[28] (Holm, Lepp, & Ringsberg, 2005) Sweden.	To explore the potential therapeutic role of a 2-month programme of structured storytelling for people with dementia. To explore the implications about teaching for nurses in dementia care.	Pilot study. N=6 PWD and n=3 paid caregivers took part. Part of larger programme (Drama Ger), with another group focusing on dance, rhythm and songs. Content analysis of facilitator reflective diary.	None noted.	
[29] (Holmes et al., 2006) UK.	To evaluate whether live or pre-recorded music is effective in treating apathy in people living with moderate to severe dementia.	Randomised placebo-controlled trial with blinded observer rater. n=32 PWD. Live interactive music, passive pre-recorded music and silence for 30 minutes. Video recording analysed every 3 mins with DCM to assess quality of engagement.	Objective assessment of positive verbal responses in patients not possible because muting sound to blind the rater also prevented rating of verbal response. Cannot infer anything about longer-term effects. Visual imagery of seeing someone play an instrument is different to seeing someone with instrument not playing it.	80% power determined, reflected need to correct for 9 comparisons including dementia impairment and conditions. Examines Category E from DCM – engagement of subject in expressive or creative activity. Attempts to deal with placebo effect. Musicians asked to play at similar volume, display similar level physical activity and to have instruments visible at all times, including during silent period. Procedures and measures the same during all interventions.

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
[30] (Houser, George, & Chinchilli, 2014) USA.	To evaluate the impact of a 6 week TimeSlips creative expression intervention on behavioural symptoms, mood, and medication use of PWD in long-term care.	Quantitative. Cluster randomized pilot study on 2 discrete care units in one nursing home. Control cohort – standard of care activity programming (n=10), intervention group – standard of care + 2 TS sessions per week for 6 weeks (n=10). Within and between group comparisons. Data collected using CareTracker. Patterns compared within and between for 8 months, without, then with.	PI facilitated the intervention – may have introduced bias. Small sample, with gender imbalance. Short intervention period. Those inputting the behavioural data not blind to the study.	Recommends larger, more diverse, multi-centered study.
[31] (Kinney & Rentz, 2005) USA.	To evaluate the wellbeing of individuals engaged in an art program for people with early to moderate dementia and to further develop an observational tool.	Observational, using Greater Cincinnati Chapter Wellbeing Observational Tool (GCCWOT). Comparison during intervention, to participation in traditional day centre activities after it. n=12 PWD across 2 sites.	'Logistic obstacles' prevented planned number of observations. Participants didn't always attend or stopped attending. Sometimes field trips followed, rather than an activity. Researchers couldn't control other activities. Small sample size and therefore couldn't make full use of original scoring of scale. Had to aggregate observations during other activities. Art program always preceded observations during other activity, participants could have been tired or otherwise affected. Training in tool needed. No way of telling whether wellbeing extends beyond the session.	Addresses identified issue that many studies did not address reliability of measures used to assess QoL. Observation of wellbeing doesn't help to show what in the process brings pleasure and engagement.
[32] (Kontos et al., 2016) Canada.	To assess effects of 12 week twice weekly elder-clowning intervention on moderate to	Mixed methods. Pre-post study involving feasibility sample of nursing home residents (n=23). CMAI.	Time-intensive methodological demands of DCM in study with 4 time points. Not all participants available at each time point and session (e.g. infectious outbreaks, participants off unit).	Offered structured training period in rating for those doing the measuring.

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
	severe BPSD in nursing home residents with dementia.	BPSD measured using NPI-NH, QoL using DCM, nursing burden of care also assessed along with secondary outcomes of occupational disruptiveness, agitation and psychiatric medication use. Study also included qualitative methods to explore aesthetic and relational components but these were not reported here.	Variation (although minimal) in dose of intervention. Intervention dose not defined. Difficulty in rating more complex neuropsychiatric symptoms over rating aggression – could have led to lack of detectable change in them. No significant reduction in aggressive and agitated behavior shown in CMAI – maybe due to differences in how each measures agitation and aggression. Societal biases about loss of self may negatively affect staff beliefs, contributing to emotional burden that improvements in residents' wellbeing does not improve. Staff training not assessed. Resident functional abilities, chronic diseases, pain etc not assessed. Small sample size, so lacking power. No control group. No generalisability. All participants from same nursing home. Reliance on established measures that do not reflect current understanding of BPSD – esp differentiation between behaviours based on their potential cause. No differentiation between needs-driven behavior and other causes, eg pain.	Further research required to better understand resident and staff level factors that might contribute to resident behaviours and nursing burden of care. Results should be seen as hypothesis generating, future larger comparative studies needed to confirm. Recommends large, powered RCT to assess clinical and cost effectiveness.
[33] (Loizeau, Kündig, & Oppikofer, 2015) Switzerland.	To assess the efficacy and feasibility of a TimeSlips (TS) creative expression intervention for people with dementia and their caregivers taking place in a museum.	Mixed methods. Small pilot study (n=4 PWD + caregivers and volunteers). Pre-post. Interviews, written questionnaires, validated and adapted scales, self-generated observation sheet, VAS to assess efficacy and feasibility. Direct questioning of PWD about subjective wellbeing.	Intervention assessed short-term effects, long-term effects not appropriate in interventions focused on PWD. Small sample size: some participants excluded because couldn't participate at least 4 times. Larger sample with control group needed. Caregiver reported diagnosis of dementia (no clinical assessment). Found no significant differences on attitude to dementia and caregiver burden scales although all reported a positive change in attitude to dementia.	Recommends future interventions focus on 'the moment', with measurements immediately post-intervention. Also, investigation of interactions with various stages of dementia, assessment of long-term effect, and continued involvement of PWD and caregivers.

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
			Relationships between variables might have been under or over-estimated due to use of only two subscales of validated scales. Confounding variables may have influenced results (e.g. progression of dementia). More robust observational tool needed. Post-intervention follow-up would have been useful.	Qual showed PWD attributed raised mood to 'atmosphere', caregivers to social contacts. Future research should explore societal level impacts, and emotions concentration, creativity, social interaction. Recommends audio or video analysis. Long-term effects, such as dyadic communication. Should question PWD directly about emotions and subjective wellbeing and involve PWD and caregivers.
[34] (MacPherson et al., 2009) Australia.	To evaluate the effects of a 6-week art discussion programme in an art gallery for people with dementia.	Mixed methods. Sessions filmed. Participants were from community (n=15), residential care (n=8). CDR used to assess impairment, another validated instrument o assess difficulties in caring. Sessions filmed; independent raters observed and coded engagement and activity. Wks 1 and 5 coded using time-sampling methods. Focus groups. Grounded theory analysis of transcripts.	Small sample, making it necessary to collapse the rich observational data collected into less fine-grained categories. Couldn't capture everything in these (notes anecdotal memorable moments). Some participants excluded from analysis because didn't attend first session. Did not formally measure baseline and post behavior in participants, family members and staff. Attempted focus groups 2-3 weeks after with residential care group: not all participants capable of conversation, one had clear memory of programme, two some recollection when prompted, one flashes of memory, two no memory at all. Content elicited not included. Focus group 6 week after with community participants all had good ability to remember.	Carers reported no lasting changes – you do it for the moment. Cannot comment on longer-term effects. Notes difficulty of assessing 'internal states' of PWD where direct measureof subjective experience not possible. Recruitment, local Alzheimer's Association - for those who might benefit. Recommends future evaluations include larger samples and wait-list control. Suggests monitoring before and after gallery visit.
[35] (Maguire et al., 2015)	To assess the effect of a 4-month programme of 3 sessions per week of active singing, on	Quantitative. Before and after cognitive testing (MMSE, Clock Design and	Norovirus outbreak led to quarantine period in middle of study, meaning no access to participants during that time.	

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
USA.	measures of cognition and life satisfaction for people with dementia in an assisted living facility.	Satisfaction with Life Scale). n=45 participants, with rating throughout by RAs. Each group contained singers and listeners (categorised according to how much they engaged).		
[36] (McCabe Greasley-Adams, & Goodson, 2015) UK.	To reflect on the success, from the perspective of those involved, of a creative musical project led by Scottish Opera which included PWD and carers in co-creation. To make proposals for developing and improving similar projects in future.	Independent evaluation, with a participatory approach. Engaged directly with PWD and carers. In depth interviews following end of project. Qualitative thematic analysis.	Participants self-selecting. Small exploratory evaluation – findings not generalizable. Reliance on interviews so findings reliant on subjective reflection and recall of participants Didn't include data collection from start of project. Participant observation would have added another perspective and more depth, and enhanced validity through triangulation with interview data. Evaluation could have been more user-led to guide researchers to more meaningful outcomes.	Authors suggest project raises issue of which outcomes projects should focus on – measurable outcomes to 'prove' efficacy or meaningful outcomes for PWD and carers? More user-led evaluation would guide researchers to meaningful outcomes. Evaluation engaged with only 27% of all those involved overall, and 44% of those taking part in the performances. Authors note rigorous research and evaluation findings can help support funding for interventions in future.
[37] (McDermott, Orrell, & Ridder, 2014) UK.	To explore how and why PWD find music beneficial for their wellbeing. To explore the meaning and value of music from perspective of PWD. To develop a theoretical model for music in dementia.	Qualitative. Focus groups and interviews with care home residents with dementia and families, day hospital clients with dementia, care home staff and music therapists. 2 care homes. n=53 (care home resident n=12; day hospital clients n= 4, family carers n= 15, care home staff n=14, music therapists n=8). 6	Many PWD interviewed moderately to severely cognitive impaired; speech often unclear or easily misinterpreted. So, joint focus groups for residents and family members held. Comments from majority of PWD in study not 'concrete' enough to be used as quotes in reporting. Focus groups with staff and residents organized in consultation with managers to minimise disruption. Arranging a mutually convenient time for family carers to attend led to more interviews being conducted than had originally been planned.	Developed psychosocial model of music in dementia. Notes previous studies have made limited attempts to link findings with theoretical frameworks and identifies need to go beyond summarizing study outcomes/ Questions whether standardized measures might create

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
		focus groups, 17 interviews. Thematic analysis.	Some individual interviews, of day hospital clients, based on variations of severity and ability to articulate thoughts. Data collection and analysis focused on specific aim and therefore some aspects of meaningfulness of music might not have been captured. Author was known to many of participants and this may have affected their responses. Not all recordings transcribed for analysis. Sample didn't include early-mid stage dementia because of restrictions of ethics and time limitation of project. Model developed not empirically tested or validated with others.	standardised inputs and fail to capture unexpected elements. Notes scarcity of high quality RCTs with people with mild / moderate dementia living in community.
[38] (McIntyre & Cole, 2008) Canada.	To provide commentary on a spoken-word performance created from data gathered from family caregivers about their experiences of caring for someone with Alzheimer's disease.	Performative.	None reported.	
[39] (Musella et al., 2009) Italy.	To assess the effectiveness of a weekly visual art and discussion programme in improving psychological health and therefore communication and life quality of people with Alzheimer's disease.	Mixed methods. n=10 people, most with mild to moderate AD, one severe. Observation of visual arts and discussion. MMSE and neuroropsych tests to baseline. Measured Mood State Test. Interviews with caregivers.	None reported.	
[40] (Narme et al., 2014)	To assess effects of music intervention versus cooking for PWD, in emotional, cognitive,	Single center RCT, with n=48 PWD. Compared music with cooking intervention.	11 of initial group of 48 did not complete whole intervention due to refusal, health or death.	Caregivers completed mood assessments (blind to intervention group).

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
France.	functional and behavioural domains, and also effects on professional caregiver distress.	Blind assessors evaluated participants before, during and after intervention (up to 4 weeks post) using videoed interviews. Discourse content measured and analysed for e.g. positive and negative words and expression of positive and negative emotions. Analysis featured within and between group comparisons.	Small sample size and non-normally distributed data – therefore non-parametric tests. Differences in demographic or clinical variables could have been adjusted for in larger sample. Previous study had shown more benefit to music intervention, but this suggests this might have been the result of implicit associations between the activity and the therapist facilitator – assessment was not blind to the intervention. No account taken of individual preferences for one intervention or other. Same facilitator delivered both and might have preferred delivering one or other.	Cooking chosen as control to test hypothesis that music would have an effect over and above the non-specific. Authors recommend a no-contact third group would show whether positive changes result from increasing familiarity with the assessor or should be attributed to placebo or natural variations.
[41] (Osman, Tischler, & Schneider, 2016) UK.	To explore the impact of group singing activities (Singing for the Brain) on people with dementia and their carers.	Qualitative. Semi-structured interviews with PWD and carers. 10 interviews with n=20 participants. Thematic analysis.	Participant numbers noted as 'relatively low' but authors note this is 'acceptable' in qualresearch focusing on understanding the knowledge and experience of a particular group. Carers tended to dominate discussions because PWD were often unable to recall sessions. All participants white British; need to evaluate activity taking place in different regions and capture experience of a more diverse population.	Little known about participants. Two versions of info sheet prepared – 1 for PWD, 1 for carers. Lead author participated in SftB sessions throughout and was familiar with interviewees. Efforts made to engage both PWD and carer in interview. Suggestions further research: standardised tests of memory, impact of different facilitators.
[42] (Palo-Bengtsson, Winblad, & Ekman, 1998) Sweden.	To explore how PWD function in social dance sessions in order to understand why social dance should be used in nursing homes.	Qualitative content analysis of videotaped dance sessions involving n=6 PWD. Adapted GBS scale for analysis of disabilities and symptoms before and after dancing (sub scales measure motor,	Participants chosen based on whether they usually took part in social dance session and enjoyed it. Frail health meant researcher couldn't know whether participant would take part until day of session. Not all took part in same number of sessions. Participants chosen after viewing	Authors suggest it was not meaningful to compare numerical scores from GBS scale and non-numerical analysis of dancing.

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
		intellectual, emotional functions and symptoms characteristic of dementia).	videotapes of sessions. Carers retrospectively assessed participants' condition.	
[43] (Palo-Bengtsson & Ekman, 2000) Sweden.	To explore the phenomenon of dance events as a caregiver intervention for people with dementia resident in a nursing home from perspective of caregivers.	Phenomenological. Qualitative. Unstructured interviews with n=7 caregivers. Interviewees shown a video recorded in the same nursing home, to support narratives.	None reported, but reflexivity discussion.	In reflexive discussion, notes need to consider what becomes data and accept the possibility of a multiplicity of responses resulting from it. Showing the video may have created a particular set of responses.
[44] (Palo-Bengtsson & Ekman, 2000) Sweden.	To examine the emotional response of people with dementia resident in a nursing home to social dancing and walks in order to understand the feasibility, popularity and meaning of activities from the perspective of the patient.	Qualitative. Descriptive phenomenological (Husserl). Videos of one social dance event and one walks for each participant, analysed. n=6 PWD.	None specifically reported, although reflexivity discussed.	Addresses issue that little research is carried out to study how PWD react emotionally to caregiver interventions. Bracketing of researchers' own beliefs, ideas, theories and personal and theoretical knowledge No intention to classify or measure emotions or to pinpoint them on an intensity scale.
[45] (Petrescu, MacFarlane, & Ranzijn, 2014) Australia.	To explore the effect of poetry writing workshops on psychological function in people with early stage dementia.	Qualitative. Exploratory. Structured interviews with n=4 participants.	Small sample – difficulties recruiting. Although study was widely advertised to carers through Alzheimer's Australia network, only 4 took part.	Notes few studies of use of poetry in PWD in community.
[46] (Phillips, Reid-Arndt, & Pak, 2010) USA.	To test the effect of a 6 week twice-weekly TimeSlips creative expression intervention on communication, neuro-psych symptoms and QoL of long term care residents with dementia.	Quasi experimental, 2 group, repeated measures comparing TS with usual care. Hypothesis testing pilot study. Baseline, and post intervention measures taken at weeks 7 and 10. CSD, NPI,	6 different facilities. Individual facilities assigned to either intervention or non-intervention. Concerns about potential imbalance of group sizes led to facilities not being assigned randomly. Initial sample n = 56. 178 invited to participate, 123 family members etc agreed to contact, 26 participants refused. 42 of remaining did not meet	Because a pilot study, sample size not determined by power considerations. Staff participants each received $20 for each data collection visit.

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
		Functional assessment, OERS. Participants n=28 PWD across intervention and control (no intervention) group. Selected from 4 nursing homes and two assisted living facilities.	inclusion criteria. During study, 5 further exited. Participants chose whether or not to attend on any given day. Small group sizes. Ethical processes, privacy concerns at homes led to 2 distributing recruitment materials themselves. If no court appointed legal guardian, waiver of consent process approved. This was the case for most participants. No participants able to make independent decisions about participation. Intervention delivered in varied environments, eg, private group meeting room and shared living space in each facility. Also, variety of settings and range of cognitive impairment, e.g. MMSE higher in assisted living than for nursing homes. Sample lacked ethnic and racial diversity. OERS inter-rater reliability declined at study mid-point, improved after retraining. Observations taken during TS and mealtimes so as to minimize differences, however, both activities naturally stimulate alertness. Participants not randomly assigned to treatment or control within same facility, couldn't isolate the active ingredient of TS from unintended effects.	MMSE conducted only at baseline and week 10 to limit participant burden. Outcomes appropriate for participants with more advanced dementia (NPI and communication) might not be appropriate for participants with milder disease.
[47] (Rentz, 2002) USA.	To provide feedback about use of a program of art-making and its effect on the sense of wellbeing and self-esteem of people with dementia taking part. To create and test a wellbeing observational tool (GCCWOT)	Observational framework. Outcomes-based evaluation, using Lawton's framework conceptualising wellbeing. 6 sites, at each a staff member observed and evaluated one participant for 60 minute session. n=41 artists participants evaluated at 6 sites.	Applied new tool. Staff did the observation and might have been biased. Multiple raters used, inter-rater reliability not assessed. Highlighted issues around the instrument, including vagueness and lack of clarity. Demographic data not recorded. Difficult to ascertain whether outcomes resulted from art activity, other intervention, or social interaction.	Could consider conducting serial evaluations to avoid errors due to artist participant fatigue or anxiety.

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
48 (Roe et al., 2016) UK.	To evaluate an art program in a gallery for older people living in a care home, including those with dementia.	Mixed methods. Observational framework.	Small sample size. Few participants attended all or most of the sessions. Only one programme, short time frame. No before and after data.	Programme wasn't specifically for PWD, but focus was on inclusivity and rights of older people to engage, regardless of diagnostic label. Project had previously attempted to use Outcomes Star, discontinued because feedback from care staff was that it was not appropriate.
49 (Sakamoto, Ando, & Tsutou, 2013) Japan.	To assess the effects of weekly individualised sessions of a passive and interactive music intervention on people with severe dementia.	Quantitative. RCT design. Passive and interactive individualized music intervention, with a no-music control. n=39 participants. Short-term effects measured via emotional response and stress levels (autonomic nerve index and Faces Scale). Long-term BPSD effects evaluated using BEHAVE-AD Rating scale, 2 weeks prior, after final session and 3 weeks following. Videoed participants during music intervention and quantified minutes of particular kinds of behaviour.	In control and passive groups, observation from a distance. Mixed facilitators to avoid potential effects of individual facilitator. Primary authors not involved in intervention or evaluation. Interventions all conducted 10-11am, but no testing as to whether this was the optimal time. No comparison to group intervention. Small sample size. Had to tightly control for confounding factors and exclude comorbidities. Only included people with severe dementia and therefore results cannot be generalized to people with different severities.	Participants examined individually in familiar room to minimise confounding effects of testing environment. Music selected for individual. Trained facilitators. No participants dropped out. Raters blinded. Suggests varying length of interventions to determine most effective duration, and including longer follow-up.
50 Särkämö et al., 2014) Finland.	To determine efficacy of music intervention based on 10 week programme. coaching intervention with caregivers of people with dementia using either singing or music listening	Single blind RCT. PWD-caregiver dyads. Mild-moderate dementia. N=89. Singing n=30, listening n=29, usual care control n=30. Extensive neuropsych assessment, cognitive tests,	Did not focus on any particular dementia type – limiting specific conclusions that can be drawn, potentially this means the sample is representative of wider PWD population and perhaps more generalizable.	Follow up sessions might have encouraged caregivers to keep up activities at home and have helped maintain positive effects.

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
	regularly as part of everyday care.	mood and QoL before, after and 6 months later. Psych wellbeing of family members assessed as well as burden (GHQ and ZBI). All assessments blinded. Short semi-structured telephone interview at follow up (Likert scale answers). Coaching included singing/listening familiar songs, exercises, rhythm (singing); reminiscence and discussion (listening).	Did not focus on very early stage dementia (MCI) during which the intervention might have been most beneficial. Intervention period was short because of funding and scheduling issues.	
[51] (Sauer et al., 2016) USA.	To explore the impact on wellbeing of participation in a person-centered and intergenerational arts activity programme (Opening Minds through Art) in which a PWD and a student volunteer are paired to complete the art-making activity.	Quantitative. Exploratory study. OMA compared with conventional visual arts activity. Video of n=38 PWD doing OMA activity at 3 different long-term care facilities. Sub-sample (n=10) videoed during traditional visual arts activities. GCCWOT (modified).	Data aggregated. Original intention was to video each participant equal number of times, but couldn't control who participated (residents chose). This meant unequal numbers of videos. For those involved in more than one, data were averaged for each 5 min observational period, resulting in one data record per participant reflecting average across sessions. Small sample size. No demographic or health-related descriptors of sample (eg severity or type). Cannot make claims about generalizability. Information about potential confounding variables not accessible. No test-retest reliability. No randomization, and inability to randomly select who would be observed across both OMA and traditional arts activities. Care setting didn't permit exclusion of residents from OMA participation. Traditional arts activities didn't include 1:1 partner component and therefore can't say whether	Authors note need to be sensitive to the cultural environment of the long-term care setting. Attention to personhood and wellbeing, rather than deficit model that focuses on cognitive failure. Authors note many programs focus on reducing BPSD rather than effect on QoL. Adapted GCCWOT to measure both positive and negative components of wellbeing. Included 25 indicators, 6 more than the original 19.

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
			differences due to interactions of residents and volunteers or to type of arts activity. No elicitation of subjective experiences from PWD themselves.	
[52] (Sixsmith & Gibson, 2007) UK.	To examine the role of music and music-related activities in the everyday lives of people with dementia.	Qualitative. Interviews with n=26 PWD and carers living in own homes or in residential care. Observations of home and interactions with carer. Care staff provide information where relevant.	Dementia diagnosis was that of service provider, not necessarily clinical. Preliminary, exploratory research.	Sampling approach emphasized diversity of personal attributes, life contexts and residential environments.
[53] (Stevens, 2012) Australia.	To evaluate the effects of a programme of 8 2-hour stand-up and improvisation workshops for people with early stage dementia, and a final performance.	Mixed methods. Qualitative. Telephone interviews with participants (n=6), carers (n=6). Observational field notes of 3 of the workshops, and final performance. Thematic analysis.	Telephone interviews chosen because participants could choose convenient times, be in familiar environment and fitted with busy schedules. However, no participants could remember having attended programme or performed or attended respite care at all – carers said this was standard. Carers provided majority of interview data. Limited triangulation. Small sample size.	Participant group already meeting regularly before intervention. Retrospective telephone survey not an appropriate method - should have done something 'in the moment'.
[54] (Sung et al., 2006) Taiwan.	To evaluate the effects of a 4 week group music with movement intervention on the occurrence of agitated behaviours of institutionalized older people with dementia.	RCT. Control was usual care. Modified Cohen-Mansfield Agitation Inventory. n=36 participants across both groups.	Small sample size, drawn from one care facility. Limited generalizability. No follow-up for longer term effects. Behaviour assessed by nursing staff. No blinding.	In future use music based on individual preferences. Compare to pharmacological interventions. Investigate confounding variables such as medication and physical restraints. Examine correlations between severity and intervention effects. Follow-up assessments. Evaluate in different settings.
[55] (Sung et al., 2012) Taiwan.	To evaluate the effects on anxiety and agitation of insitutionalised older adults with	RCT. Repeated measures. N=60 participants across both experimental and control	Three participants did not complete (hospitalization and relocation). Low occurrence of agitated behaviours in both groups at baseline. Sample	Co-existence of anxiety and other BPSD makes assessment complicated. Scales for anxiety

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
	dementia, of a 30 minute group music and movement intervention using percussion instruments and familiar music, delivered twice weekly over 6 weeks.	(usual care, no intervention). RAID and Cohen-Mansfield Agitation inventory at baseline, and weeks 4 and 6.	drawn from one facility. Blinding not possible. Lack of comprehensive tool for measuring anxiety and agitation – more development and refinement of tools needed. Extraneous variables may have affected. Longer follow up needed.	in dementia vary and can overlap with depression scales. Suggests reductions in anxiety and agitation in experimental group may have had knock-on effects for control group in same institution. Future: use participants from same residential area as experimental / control, with crossover design.
[56] (Swinnen, 2016) USA.	To explore the use of poetry interventions in dementia care.	Qualitative. Ethnographic participant observation of 19 poetry interventions. Audio recordings to supplement thick description and verbatim transcription. Fieldnotes and interviews. Poems used in intervention, video and photographs. Content analysis.	None reported.	Author suggests ethnography reflects voices of PWD, allows researcher to familiarize with repertoire of individual responses of individual participants. Suggests that you can only answer question of how poetry works in this setting, by looking at characteristics of intervention as a live participatory event.
[57] (Ullán et al., 2013) Spain.	To describe and explore the effects of an artistic educational programme based on photographic cyanotype techniques, for older people with early stage dementia.	Exploratory qualitative study. Participant observation. Engagement assessed by educators (study specific). n=21 participants with mild/moderate dementia. Focus groups with participants, educators and professional carers. Video recordings of interviews reviewed by two assessors.	Logistics or health prevented 6 participants participating in focus groups. Small sample size. Not all participants attended all sessions and sessions included varied group sizes. No detail of dementia type. Only conducted in one setting. Can't draw conclusions about longitudinal effects.	Suggests is not possible to evaluate PWD artistic activities on just one level. Focus groups included participants and the educators who had directed the group. Important to view art activity involving PWD as a 'performance context

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
				unconditioned by dementia' (p438).
[58] (Unadkat, Camic, & Vella-Burrows, 2017) UK.	To understand how group singing benefits people with dementia and their partners.	Qualitative. Interviews with n=17 couples (n=34 participants in total) taking part in different singing groups. Grounded theory, informed by critical realism.	Caregiver's views may be overly represented because of better cognitive and verbal abilities. Previous relationship quality, interpersonal style, resilience etc not addressed. Findings specific to participants taking part in study. Researcher assumptions cannot be completely eliminated. Potential self-selection bias, as participants volunteered for group singing. Only heterosexual couples.	Longitudinal research needed in future to explore links between creativity and relationship resilience. Bracketing used to examine researcher assumptions. Scrutiny by second authors. Comparison study between reminiscence singing interventions and singing including new learning may yield valuable results. Also compare other meaningful couple activities Sample broader range of ethnic groups.
[59] (Ward & Parkes, 2017) UK.	To evaluate a group singing activity (Singing for the Brain) conducted with people with a learning disability, memory problems or a dementia in care centres.	Mixed methods. Observation of participants (n=3), patient and staff feedback. Comparisons with regular sessions for PWD. Questionnaire developed for project (n=15). Interviews with staff (n=6), stakeholders (n=4). Thematic analysis.	Positive wellbeing, mood and memory benefits reported anecdotally, but evaluation not set up to comment on these.	No comparison with other sessions run at the day centres. Further research required to investigate longer-term impacts.
[60] (Young et al., 2015) UK.	To assess the impact of an 8-week programme of art-making and art-viewing group activities for people with dementia in an art gallery on verbal fluency and memory in PWD.	Quantitative content analysis of audio-recordings of sessions. PW early/mid D (n=13). Opportunity sampling from two groups. Codebook developed in previous study adapted and used in analysis.	Self selected opportunity sample includes people with prior interest. No control group and therefore can't separate intervention specific factors from other variables. One audio recording failed. Analysis of each utterance in random order might have decreased impact of any possible bias to	Provides support for future controlled studies. Quant content analysis as measure leaves participants unaware of analysis – reducing possibility of measurement confounding data. Allows quant exploration of qual data

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
			coding – unlikely however to have been significant because data doesn't show clear trends. Some sessions may have been 'more demanding' for participants, affecting observation that there was a change over time. Data from art-making and art-viewing presented separately to consider differences – makes data vulnerable to recency effect as art-making always followed art viewing – participants were perhaps more relaxed and confident or perhaps more tired. Intervention took place in public gallery but this meant no video and it was not always possible to identify individual participants from audio – therefore participant data grouped. Analysis at group rather than individual level, and therefore no inferential statistics possible. Might have skewed data by more or less vocal participants. Differences between groups not addressed. Data not coded immediately after sessions, which might have allowed individual identification. Measurement of lifetime memory reporting – participants could be making it up – no way to assess.	Group rather than individual data analysed – unusual for content analysis. Allowed exploration in naturalistic environment, no loss of essence of gallery experience.
Included studies from update searches 2018				
[61] (Basting, 2018) USA.	To reflect on a drama-based project for people with dementia living within a care environment, including some reporting on an evaluation of the project.	Mixed methods. Pre-post surveys to capture changes in attitude towards aging and PWD. Focus groups and interviews.	'Creative community of care' (CCC) projects are challenging to evaluate because they can evolve, shift midstream, use variety of facilitators and multiple modes of engagement, and participants come and go.	CCC projects viewed as open or complex systems in which families, staff and volunteers engage as equal partners. Care home seen as complex system itself.
[62] (Belver et al., 2018) (Belver *et al.*, 2018)	To describe the design, development and evaluation of a programme of artistic education	Qualitative. Participant observation of n=12 PWD. Researchers complete field	No control group of participants without dementia. Not possible to comment on long-term effects.	

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
Spain.	activities, based on visits to the Prado Museum, for people with dementia. To assess whether it was an appropriate activity.	journal. Thematic analysis. Group discussion with PWD, art educators (n=4) and caregivers (n=6).		
[63] (Boersma et al., 2017) Netherlands.	To evaluate the implementation of the Veder creative communication contact method (using skills from original method as part of daily care). To explore the facilitators and barriers to its implementation in natural contexts.	Process analysis. Multiple case study design. Qualitative. Nursing home ward as unit of analysis (n=6). Focus groups with caregivers (n=42). Interviews with stakeholders (n=12). Thematic analysis, using RE-AIM framework.	Use of standard protocol may have meant some themes only briefly discussed, and this became obvious in analysis process. No involvement of PWD or informal caregivers Variation in settings, and in time for implementation of intervention. Can't generalize. Interview methods provided perceived barriers and facilitators, observation would have supported and complemented them.	Strengths included: same researcher using same protocol conducted all stakeholder interviews; used predetermined analysis scheme based on RE-AIM; independent researchers checked reliability; no dependency between Theater Veder and researchers. Increased awareness of importance of implementation, adoption and sustainability. Sub-optimal implementation can lead to poor outcomes in effect studies (implementation error). Different levels of stakeholders might have different expectations and interests.
[64] (Broome, Dening, & Schneider, 2017) UK.	To explore the factors affecting the successful facilitation of an arts programme in residential care from the perspectives of artists involved. To identify barriers and facilitators.	Exploratory qualitative. Views of artists, n=32 captured in reflective diaries. Thematic analysis.	Small scale study, not possible to generalize. Only one approach and data collection method used. Response bias – artists were commissioned by arts organisation to whom they were reporting in the diaries.	Recommends future research includes investigation of views of people with dementia and care professionals involved. Could also explore the impact of different arts forms.
[65] (Broome, Dening, & Hill, 2018) UK.	To describe two case studies of arts interventions in UK care settings, involving visual arts and dance movement activities for people with dementia.	Qualitative case studies. Interviews (artists, n=2; managers, n=2, activity co-ordinators n=2), reflective diaries (n=19) and narrative	Case studies selected because fitted with programme timing. Small scale study with limited transferability. Lacks participants' accounts.	Practitioners faced challenge of identifying purpose and expectations of the intervention, while at the same time considering the working

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
		monitoring reports (n=10). Thematic analysis.		demands of a care setting. No investigation of differences between settings.
[66] (Campbell et al., 2017) UK.	To evaluate and explore the experiences of those taking part in a 10-week music programme for people with dementia in one care home, conducted by members of a renowned chamber orchestra.	Exploratory qualitative evaluation. Interviews with stakeholders (19 interviews, n=11 participants). Thematic analysis. Researcher also attended debriefing sessions as sensitizing data. Activity worker diaries. Sample included music therapist involved, musicians, organisational team, activity workers and care home manager.	Request to conduct evaluation received only 2 months before project began; PWD lacking capacity could not be included because of lengthy requirements of ethics process. Family members approached but none consented. No independent measures of dementia diagnosis or level of severity. Only 2/3 activity workers gave permission for diaries to be used No longitudinal perspective.	Authors reflect on difficulties of measuring intangible aspects of creative process, including atmosphere generated, how to record 'shared identity' moments, capture embodied meanings and expressions, and operationalize concept of 'embodied selfhood'. Sample weighted towards musicians and organization rather than participants. Video, photographs and other multi-media used to document – providing 'visual and auditory testimony'. 'Memory jogger' photographs used to remind residents about the visits. Notes potential for longitudinal application of wellbeing scales alongside real-time and in the moment experiences. Also potential for video work and sensory methodologies.
[67] (Daykin et al., 2018) UK	To assess the effects of a 10 week period of weekly participatory music sessions taking place in an acute hospital elderly care service, on	Pilot mixed methods study. Exploratory sequential mixed methods design. Descriptive quantitative ward level data for equivalent time periods, one with and without music,	Difficulties of research within acute hospital settings – busy staff, patients short term and receiving different numbers of sessions. Differences in ward environments time periods assessed could have resulted from e.g. average age of patients, combinations of diagnoses, comorbidities etc.	No distinction between different types of dementia at recruitment because they don't influence clinical care provided in acute setting.

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
	wellbeing and on the ward environment.	n=85 patients with dementia diagnosis. Observations (ArtsObs) and detailed field notes. Interviews and focus groups. 9 brief conversational interviews with participants immediately or short time after session with primary n=38 patients and n=12 staff.	Interviews with participants difficult and typically consisted of short, fragmented answers, needed frequent prompts and often struggled to remember. Hearing impairments and other physical difficulties contributed challenges. Some declined interview: feeling anxious, needed to be seen by doctor, didn't want to miss visitors. Beyond scope to seek detailed data in each time period on patient characteristics, therefore no control for confounding variables. No inferential analysis possible. No matching or control group. No before and after measurement.	Clinical nurse specialist handled recruitment and took consent. Only 6 out of 12 staff took part in final focus group. Music session had to fit into hospital routines, regularly interrupted. Further research needed to understand personal, interpersonal and social factors mediating wellbeing outcomes from music and arts interventions.
[68] (Eades, Lord, & Cooper, 2016) UK.	To evaluate an outreach programme designed to engage socially isolated people with dementia by offering them opportunities to engage in and enjoy cultural activities with artists in their own homes.	Qualitative. Interviews with n=6 PWD, n=4 artists, n=3 befrienders. Socio-demographic info questionnaires.	No control. Can't tell whether enjoyment and engagement reported are result of visit and companionship or arts activities.	Open-ended, simple, repeated questions in interview topic guide for PWD. Festival in a Box archive used as a visual prompt.
[69] (Evans, Garabedian, & Bray, 2019) UK	To evaluate a music-based group reminiscence programme aimed at improving wellbeing of PWD and their carers. To make recommendations for future practice.	Mixed methods. Pre and post. Familiarisation with programme, observation using CEAA tool. Descriptive statistics. Process consent at each point of contact. Documentation of programme. Feedback collected from volunteers. Focus groups with carers and volunteers. Interviews with dyads at 3 month follow up.	'Go along' interviews with PWD planned but it was difficult to find right moment to be alone with participants and that participants had limited communication skills. Potential confounding factors for CEAA scores. Comparison across sessions not reliable because tool doesn't account for individual's current mood, theme of session, affinity with music on the day, number of volunteers, or level of interaction between volunteer and participant. CEAA provides only one measurement for each participant across intervention, doesn't allow	Dyad interviews – captures particular perspective. Need a tool that is suited to a specific psychosocial intervention. Authors raise question of how it is possible to fully involve PWD in evaluation of complex interventions.

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
			identification of subtle details of experiences and reactions of participants.	
[70] (Hendriks et al., 2018) Netherlands.	To evaluate the implementation of an interactive art program for people with dementia and their caregivers in 12 museums. To explore the organizational impact on the museums, and on attitudes towards dementia of museum staff.	Mixed methods. Qualitative, multiple case study. Interviews with n=23 stakeholders to identify facilitators and barriers to implementation. Meildand framework used as coding schedule. Attitudes of n=176 staff measured using ADQ (pre / post one group design)	Exploratory. May not be applicable for all museums and in other countries. Interviews mainly held in starting phase, and thus facilitators and barriers identified partly based on expectations of stakeholders. Cannot generalize. Uncontrolled.	
[71] (Johnson et al., 2017) UK.	To compare the impact of two museum-based group activities (including art-viewing and object handling) and a social activity on the subjective wellbeing of people with dementia and their caregivers.	Quasi-experimental cross-over design. Repeated measures. N=66 PWD and caregivers participated in museum object handling, followed by a refreshment break, and then art-viewing in small groups. VAS scales used to rate subjective wellbeing pre and post each activity. Brief feedback questionnaire.	Convenience sampling. Cannot generalize. Not known whether participants had existing interest in art. Gender imbalance. Of dyads who initially registered interest in participating, 9 were unable to attend due to illness, 15 changed their mind about participating without giving a reason, 10 found other life events taking precedence. Dimensions captured by VAS limited in scope and comprehensiveness. Meaningfulness of results dependent on whether participants understood concepts, although authors note few expressed a great deal of difficulty. May have been insufficiently powered.	Same facilitator for all sessions. VAS chosen so as not to burden participants with lengthy/cognitively challenging measures. Break may have been of insufficient duration – making it equal length to other interventions would have enabled more robust comparison. Recommends future research include multiple sessions over longer time period. Could add measures to assess cognitive domains. Three-armed RCT would provide more robust evidence. Observational methods, such as video to code aspects of physical engagement with objects alongside VAS. Broaden research to include

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
				people with more severe dementia.
[72] (Keating, Cole, & Grant, 2018) UK.	To evaluate the effectiveness in improving overall quality of life of group reminiscence arts sessions for PWD, conducted over 24 weeks and across 6 care homes.	Comparative and time series design. Comparison with 6 care homes not receiving intervention (usual care). DCM used to measure positive behaviours and rate QoL before, during and after sessions at 3 weekly intervals, with a 3-month post-intervention evaluation to assess sustained effect. n=75 PWD. at start.	Small scale pilot study. Observations limited to everyday activities. Not possible to make conclusive statements about impact on QoL. Randomising selection of care homes not possible due to uncertainties around availability within time frame of project. Control groups showed statistically significant more positive behaviours than those involved in the intervention. May be because other structured activities were taking place in those care homes (e.g., some homes taking part in a dementia trial on person-centred practice). Participants in these homes had higher mood/engagement scores at baseline and scores remained stable during study. No group matching. Possible DCM is not good at capturing long-term benefits of group reminiscence arts sessions, better at capturing micro changes during an activity. Measured changes at group level – may have been better to measure change at individual level. Study didn't capture and explore views of PWD. Recruitment may have introduced bias - staff may have been preferential in selecting who took part. Couldn't assess whether severity of dementia was predictor for improved well being because of a non-specific dementia diagnosis for over half participants in intervention group. Couldn't assess for differences between types of care homes. Activities very variable in content and length, not possible to identify which elements were beneficial.	No other non-reminiscence or structured arts activities were taking place in intervention homes. QoL defined in terms advocated by WHO – includes context in which people live, relationships with others and with environment. Residents with dementia identified by care staff. Challenge of comparing non-standard activities with iterative standardized activities More work needed in measuring longterm effects and sustaining them. Would have been beneficial to have some kind of self-report or other subjective measure in addition to DCM. It was not possible to evaluate how far evaluation enhanced overall quality of care.

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
[73] (Kontos et al., 2017) Canada.	To examine the effects of a 12 week programme of elder-clown practice involving residents of dementia unit in a long-term care facility.	Mixed methods. Qualitative interviews with practitioners. Ethnographic observations of video-recorded interactions. Practice reflections. Other part of study was quantitative (reported elsewhere).	-	Comments on criticism that use of video may entrain the Hawthorne effect – with conduct being modified by the presence of camera. Authors habituated participants to use of the camera and suggest that it offers a 'unique epistemological perspective' and can capture everyday practice. Footage can be viewed again and in greater depth.
[74] (Longden et al., 2016) UK.	To evaluate the impact of shared reading groups for people with mild to moderate dementia resident in a care home.	Service evaluation. Cross-over wait list design. Two homes randomly assigned to read-wait, others to wait-read. Pre post measures; baseline and then monthly. QoL measured by DEMQOL-Proxy and BPSD by NPI-Q (only symptom severity). N=36 people with mild/mod dementia, and 31 datasets in final analysis. Interviews also conducted, but results of these not reported here.	Dementia not assessed formally, identified by care home managers. Small sample. Attrition – one died, two withdrew, two were included at baseline. Lack of statistical power. Potential for clustering effects. Demographics and clinical details not collected as outside primary concern of evaluating impact. Self-selecting sample. However, residents were not included if they had violent/disruptive tendencies, became distressed in group, or were unable to sit still for reasonable time. Homogeneity. Ceiling effects - very low NPI-Q scores – so these data not analysed further. No control, no investigation of mechanisms of impact.	DEMQOL Proxy used because of challenges of direct QoL assessment. Authors suggest proxy measure improves likelihood of detecting changes attributable to group rather than other situational factors. Also, that it allows insight into perceived impact of intervention for care-givers. Challenges of direct QoL assessment in dementia – difficulties resulting from communication problems, impaired insight, recall and time perception.
[75] (Low et al., 2016) Australia.	To evaluate the feasibility of a thrice weekly 16-week group dance programme for people with moderate/severe dementia	Quantitative. Pilot RCT. Dance group (intervention), music appreciation and socialization group (control).	Recruitment easy and attrition only 3. However, attendance at sessions poorer than anticipated. Timing may have impacted this – control held immediately after breakfast, residents escorted	Assistance of staff and volunteers was critical in delivering program safely.

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
	living in nursing homes with regards to recruitment, retention, assessment tools, intervention safety, attendance and engagement.	Descriptive statistics for feasibility and attendance. MMSE. SIB – neuropsych tests for severe dementia; SPPB measure assesses physical function; CGIC provided blinded assessment of clinician's view of global function before and after intervention.	there; dance held late morning, residents had to be escorted from own rooms where many were asleep. Halfway through improved attendance by serving morning tea in activity room just before session. Measures. Authors note possible ceiling effect on SIB. SPPB showed deterioration after 16 weeks but was appropriate. CGIC feasible (including video-recording), but logistically challenging to conduct interviews with staff and residents. Single site – limited generalizability.	Control might need to be reconsidered. Dance teacher led both and both included cognitive stimulation and socialization.
[76] (Mc Parland, Cutler, & Innes, 2017) Ireland.	To evaluate a pilot 10-week fortnightly group community music and movement project designed to provide social interaction, peer support, engagement, active participation and choice and control for people with dementia and their carers.	Evaluation. Ethnographic, video recordings and field notes. Informal discussions recorded at end of each session. Supported by 1:1 conversation with some participants. Analysis using NVivo	Limited time, evaluation commissioned at short notice, with limited funding. Tight deadlines for recruiting, planning evaluation and ethics processes. Detailed histories of participants not known, no prior measure of wellbeing taken, no comment can be made on improvement in wellbeing over time. Practitioner not able to get to know participants needs and preferences beforehand. Practitioner had no prior experience with PWD. Paperwork presented at first session, including separate consent documentation for PWD and carers – thought to be over-burdensome and some potential participants opted out of evaluation.	Discusses challenges and consequences of delivering and evaluating innovative project on short notice and with limited funding. Notes poor communication. More time to explore aims with the organization and the practitioner would have added clarity and focus to the intervention.
[77] (Mittelman & Papayannopoulou, 2018) USA.	To evaluate the effect of participation in a chorus for PWD and family carers.	Mixed methods. Used forms from MoMA evaluation. QoL-AD and Dem-QoL self report. Caregivers also completed several scales. GDS – dementia severity. Structured and open-text questionnaires. Focus groups (not recorded, notes only). N	Sample size too small for definitive analysis. Low baseline for depression in caregivers. No resources to conduct formal qual analysis. Positive self-image in responses of PWD at intake and follow-up may have resulted from sampling bias or tendency at early stage to deny limitations.	Reports on 6 years of running the chorus, but study itself conducted right at the start. Aim was to fine-tune intervention, obtain initial qual and quant data on potential benefits to inform future trial.

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
		= 11 dyads, 10 completed pre and post.		
[78] (Mondro et al., 2018) USA.	To describe and summarise feedback from an art-making program for people with dementia and their care partners living at home in which caregivers were given training in incorporating art into daily life for enjoyment, relaxation and personal expression.	Pilot study to describe process of designing and implementation and summarize feedback at conclusion. Qualitative survey at end. Paper based. Thematic analysis.	Recruitment challenging. 3 years to recruit 30 dyads. Not all attended all sessions. People viewed themselves as not artistic and there was resistance to committing to 8-week programs. Limited budget to advertise.	Optional respite care was offered for PWD during training programme. Suggests building partnerships with local organisations to support recruitment.
[79] (Newman et al., 2019) UK.	To explore the experience for people with dementia of taking part in visual arts enrichment activities and to examine how and whether it provided opportunities for supporting their resilience.	Qualitative. Used data from Dementia & Imagination study, care homes n=48 PWD, n=37 carers. Interviews at baseline, post and 3 months after with sub-set of participants, and separately with carers/family members. Structured notes of artists. Videos of sessions. Socio-demographic and clinical data. CDR. Structured field notes made at session.	-	Concept of resilience: term often used but applied differently, challenges for measurement and differences in understanding. Subjective. When exploring concept with people with advanced dementia, can't rely on traditional interview schedules in research into it. Only 3 PWD had capacity to give consent. Dementia symptoms vary hour to day, and over time.
[80] (Schall et al., 2018) Germany.	To assess the impact of six sessions of an art museum-based intervention for people with dementia and care partners intended to promote emotional well-being and improve quality of life museum visits followed by artistic activity.	Randomised wait list controlled study. PWD – mild/moderate (n=44) and care partners (n=44). Independent museum visits as control. Measures before first and after last museum visit; control measure immediately after	Participants recruited using newspaper ads and advice from the Alzheimer's Association. Six did not complete programme. 25 dyads took part in intervention, 19 in wait list control (some couldn't visit independently, therefore control was smaller). Self-report on some scales not possible beyond certain degree of impairment – but assessment of situational wellbeing using smiley face was possible for all.	Future research should include combination of measurement and assessment techniques. Suggests use of procedural evaluation methods such as times series analysis, perhaps alongside longitudinal design. Suggests studies should focus on social and interactive

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
		independent visits. MMSE, ADAS-Cog, GDS, QoL-AD, NPI. Self-rating emotional wellbeing smiley face scale from Questionnaire of General Habitual Wellbeing before and after each visit. Follow up with relatives of PWD c3 months post. Caregivers also evaluated each single visit on basis of specific criteria – prepared in accordance with CODEM (included open response)	Exclusion criteria included a selection bias and sample only partially met requirements for random sample. Exclusions included advanced dementia, resident in a care facility, motor impairments. Participants from more educated middle class often had more interest in art and culture from the outset.	parameters. Notes no standardized instruments exist for measuring these in art-based interventions. Recommends development and validation. Notes importance of selecting and assessing outcome variables that might be influenced by use of art in PWD (eg mood and wellbeing). Some effects were not specific to intervention, but relate to experience of museum visit. Further research needed to understand effect mechanism. Recommends pictograph scales.
[81] (Swinnen & De Medeiros, 2018) USA / Netherlands.	To gain insight into the concept of 'play' and the experience of those taking part in TimeSlips creative expression programme and participatory poetry programme.	Humanities based inquiry.		Suggest focus on measurable outcomes of health, cognition and behavior risks overlooking opportunities offered by arts to socialize, use imagination and temporarily escape everyday realities and role of 'patient'. Humanities based inquiry moves away from question 'How is it true?' to focus on complexities of experience and meaning, and double hermeneutics that recognizes role of researcher in interpretation.
[82] (Tan, 2018) Singapore.	To explore the impacts of a 6 week pilot programme of reminiscence with art-making	Qualitative. Video recordings, photography, and	Exploratory, small sample size, qualitative – therefore no generalizability.	Participants did not have previous experience of art-making.

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
	within the heritage collection at a cultural heritage museum, for people with mild to moderate dementia.	artist journal entries. Grounded theory approach.		Recommends future studies incorporate interviews with participants and seek participant feedback. Recommends mixed methods, to include quant analysis of video data.
[83] (Tyack et al., 2017) UK.	To explore the delivery of art-based intervention for PWD through a tablet touchscreen device and to assess its wellbeing impacts.	Quasi-experimental mixed methods within subjects study. N=24 PWD and caregivers. Wellbeing measure taken before and after each tablet use (QoL-AD and two adapted VAS scales measuring levels of happiness, wellness and interestedness). Interviews at end of project. Thematic analysis.	Underpowered. Perhaps with smaller and different set of planned comparisons, effects could have been generalizable. Several different statistical analyses run – increasingly possibility of Type-I errors. No control group – but uncertain what a suitable control would be. Technology may have been an intervention in itself, equivalent to art. No way of separating the two. Self-selecting motivated sample, perhaps those who were more accepting of diagnosis, and had more optimistic resilient attitudes. Not specifiying number of sessions would help determine ceiling on benefit and when.	Both quant and qual findings would be less informative in isolation. Conceptualises wellbeing as a dynamic phenomenon, a state of equilibrium that can be affected by life events or challenges and therefore can be subjectively assessed. Power analysis suggested minimum sample size of n=34 PWD. Smaller size allowed easier collection of qual data. Nuanced changes with different wellbeing subdomains showing different patterns of change.
[84] (Windle et al., 2018) UK	To evaluate the impact on quality of life, communication and perceptions of the programme, of a 12 week visual art programme for people with dementia in residential care facilities, NHS assessment units and community venues.	Mixed methods longitudinal study. Repeated measures before, during and after. N=125 PWD across 3 setting types. Residential care facilities (n=4 settings). NHS assessment units (n=2 settings), including recruitment from day care service. Community venues	Observational measures require more training, time-consuming, generate substantial data. Had to modify protocol to facilitate recruitment in NHS site, where research was challenging. Study design doesn't allow definitive conclusions about effectiveness - this would require randomization and matched comparison/control.	Program developed through theoretical investigation of contextual factors and mechanisms shaping outcomes – recognized as important for MRC evaluation implementation process. Socio-demographic characteristics and level of cognitive impairment explored

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
		(n=3 settings). Data compared to alternative activity without art. QoL (DEM-QoL self-reported and proxy). Communication and social behavior (HCS – carer completed). Wellbeing measured using adapted GCCWOT. Short questionnaire post session to ascertain individual in the moment responses – smiley face. Follow up n=63 interviews. Thematic analysis. Long term follow-up at 6 months from n=101 participants.	Sample not ethnically diverse (couldn't offer participation to those without English or Welsh language).	as potential confounders in statistical analysis. Tested whether outcomes over time differed between sites. Self-reported QoL didn't change, but proxy reported did. Strengths included mixed methods (subjective individual experience to augment or contrast quant), standardized, validated QoL measures both proxy and self-report, larger sample size with low attrition, structured observation to capture what happens during delivery; robust data analysis strategy towards comparison of results from different sites.

9.4 Literature review: Grey literature list of included studies

[1] Algar, K. (2012). Lost in Art too…? Evaluation of art sessions provided by Denbighshire County Council. Bangor: Dementia Services Development Centre, Bangor University.

[2] Baker, E. (2014). *Arts interventions in dementia care.* Unpublished Doctorate of Clinical Psychology thesis. Canterbury Christ Church University.

[3] Bardouille, C *et al.* (2013). Uplift! Autumn 2013: a pilot dance and music project for people living with dementia and their companions. London: Green Candle Dance Company.

[4] Collective Encounters & Thornton, S. (2010). *Now and then: an evaluation report.* Liverpool: Collective Encounters.

[5] Daykin, N. et al. (2016). Arts and dementia: using participatory music making to improve acute dementia care in hospital environments: an exploratory study. Winchester: University of Winchester.

[6] Eekelaar, C. (2011). *Art gallery-based interventions in dementia care.* Unpublished Doctorate of Clinical Psychology thesis. Canterbury Christ Church University.

[7] Evans et al. (2015). My Musical Memories reminiscence programme: an independent evaluation report. Worcester: University of Worcester.

[8] Franklin Gould, V. (2013). Reawakening the Mind. Evaluation of Arts 4 Dementia's London Arts Challenge in 2012: arts interventions to re-energise and inspire people in the early stages of dementia and their carers. London: Arts 4 Dementia.

[9.] Fujiwara, D. & Lawton, R. (2015). Evaluation of the Reminiscence Arts and Dementia: Impact on Quality of Life (RADIQL) programme in six care homes. London: Simetrica.

[10] Garabedian, C. (2014). 'I'd rather have music!': the effects of live and recorded music for people with dementia living in care homes, and their carers. Unpublished PhD thesis. University of Stirling.

[11] Gant, K., Hetherington, J. & Reynolds, J. (2011). *Telford arts and dementia programme.* Creative Health CIC.

[12] Gregory, S. & Windle, G. (n.d). Lost in Art Too: an evaluation of a 10 week programme of art sessions provided by Denbigh County Council for people with dementia and their carers. Bangor: Dementia Services Development Centre, Bangor University.

[13] Hara, M. (2013). *We'll meet again: music in dementia care.* Unpublished PhD thesis. University of Exeter.

[14] Harries, B., Keady, J. & Swarbrick, C. (2013). The Storybox Project: examining the role of a theatre and arts-based intervention for people with dementia. Manchester: University of Manchester.

[15] Hirst, K., Lane, R. & Mikelyte, R. (2012). Performance-based activities in dementia care: celebrating the ‘Here and Now’ together. Bright Shadow.

[16] Irving, A. (2011). Promoting wellbeing and combating isolation: arts and dementia pilot project. Newcastle: Centre for Public Policy, Northumbria University.

[17] Johnson, J. (2015). *Psychosocial interventions and museums.* Unpublished Doctorate in Clinical Psychology thesis. Canterbury Christ Church University.

[18] Kimber-McTiffen, K. (2012). Dancemind – Alzheimer’s project evaluation report. Dancemind.

[19] Lloyd, J. (2016). The role of reminiscence arts in the lives of care home residents living with dementia. Unpublished PhD thesis. Royal Holloway, University of London.

[20] London Centre for Dementia Care & Central & Cecil (2010). Over the Moon: effectiveness of using interactive drama in a dementia care setting. Central & Cecil.

[21] Minghella, E. (2016). *Dance for Life: an evaluation of the pilot programme.* Dementia Pathfinders / New Adventures ReBourne.

[22] Morphew, G. & Murdin, A. (2013). *A right old song and dance.* Arts in Hospital.

[23] Musica (2016). Music and Memories project evaluation 2015/2016. Musica.

[24] Nicholson et al. (2015). Reminiscence Arts and Dementia Care: Impact on Quality of Life, 2012-2015. A final report of the evaluation, December 2015. London: Royal Holloway, University of London.

[25] Tyack, C. (2015). *Touchscreen interventions for people with dementia.* Unpublished Doctorate in Clinical Pyschology, Canterbury Christ Church University.

[26] Vella-Burrows, T. & Wilson, T. (2016). Remember to Dance: evaluating the impact of dance activities for people in different stages of dementia. Canterbury: Canterbury Christ Church University.

[27] Waugh, J. (2015). Evaluation of arts & reminiscence activity programme in residential care homes in the Wychavon District Council area. Jenni Waugh Consulting Ltd.

[28] Whitaker, K. (2016). *Music for Life programme report: April 2015 – July 2016.* London: Wigmore Hall Learning.

[29] Young, R. (2014). *The cognitive impact of art-gallery interventions for people with dementia.* Unpublished Doctorate of Clinical Psychology thesis, Canterbury Christ Church University.

9.5 Literature review: Table of characteristics of included grey literature

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
[1] Algar (2012) UK	To explore the role of a 10 week visual arts project in addressing issues such as social isolation, confidence, communication, quality of life and wellbeing, for people with	Observation, artist notes and focus group led by artist.	None reported	

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
	dementia and their family members and carers.			
[3] Baker (2014) UK	To investigate the mechanisms of impact for arts-based interventions at two art galleries involving people with dementia and their carers in art-viewing and art-making.	Qualitative Clinical Doctor of Psychology dissertation. Grounded theory. Interviews, field notes and written communications.	Future outcomes-based research would need to explore the implication of subjects not remembering the experience. Limited sample including only people with mild/moderate dementia aged 55+. Voluntary recruitment, may have had existing interest in art.	
[2] Bardouille et al. (2013). UK	To evaluate a pilot dance and music project (10 weekly workshops delivered in a day care centre) for people living with dementia and their companions. To demonstrate improvements in wellbeing and functional abilities. To measure improvements in cognitive or physical abilities.	Mixed methods. Survey / feedback forms pre, mid and end of project. Photography and video. OT outcome measures at weeks 1,5 & 10. Monitoring data.	None reported.	Where participants couldn't complete self-evaluation themselves, carers completed on their behalf.
[4] Collective Encounters & Thornton, S. (2010) UK	To assess the success of a project using theatre to explore and platform the experiences of people with dementia and those who care for them.	Audience feedback and quotations from participants, but no details of collection or analysis. Monitoring.	None reported.	Mentions detailed evaluation plan and varied evaluation methods, but no details of these are given.
[5] Daykin et al. (2016) UK	To investigate the effects on patients, carers, staff and the ward environment of a 10 week music programme for people with dementia in an NHS hospital.	Mixed methods. Ward and level data collected during two equivalent 10 week time periods, with (A) and without (B0 music. Participants n=85; n=38 at time A; n=47 at time B. Participant observation (ArtsObs scale), semi-structured interviews and	Many factors could have contributed to differences in ward environment between two time periods. Missing data. Variations in group size and composition. Effect of underlying health conditions. Responses may have been affected by medication changes. Participants found it difficult to participate in interviews. No matching or control group.	Consent should be shared responsibility between researchers and clinical nurse specialist. ArtsObs tool and observation invaluable because of difficulties with interview – but time consuming. Staff reflective diaries used for personal reflection / discussion,

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
		focus groups. Action research component with staff.	Time commitment required of nurse specialists manually screening medical notes and records and supporting recruitment. Secure storage facility required on site for study documents, including consent forms. Clinical workloads took priority. Pragmatic sample size not achieved. Inferential analysis not possible. Research governance approval delayed, and delays in obtaining consent delayed starting data collection. Consent process was onerous and time consuming for clinical nurse specialist doing it.	not reported on because confidential. Recommends:: representative sample and increase number of patients and number of wards; detailed data on ward characteristics before and after detailed patient data on factors that might affect results; within subjects research study to minimize measurement error from population variation (although consider impact on resources); reduce time points; ensure valid indicators of patient improvement; buy-in from senior staff and engagement of care staff; clarify responsibility for obtaining consent; mixed methods and longitudinal study; include ward-based observations; ensure sufficient funding.
[6] Eekelaar (2011) UK	To examine changes in cognition, episodic memory and verbal fluency in people with dementia taking part in a 3 week arts-based intervention at a gallery.	Clinical Doctor of Psychology dissertation. Mixed methods exploratory study. Audio recordings. Content and thematic analysis. N=6 PWD, and n=6 family carers.	No causal link possible. Particular subtypes of dementia might respond differently. Recruitment may have been biased to those with existing interest in art. Impractical to record speech during art-viewing stage because sessions took place in public areas. Participants felt 3 weeks not long enough to acclimatize to group. Short time frame: couldn’t use standardized psych tests, plus difficult to administer, varying reliability and validity. Lengthy questionnaires difficult for PWD.	Add control group and lengthen number of sessions. Comparison with similar activity in a different setting, recommended.

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
[7] Evans et al (2015) UK	To explore the impact of a musical reminiscence programme on people with dementia.	Mixed methods. Focus groups, interviews, non-participant observation using Creative Expressive Activities Assessment (CEEA) tool, participant observation by volunteers. Thematic analysis.	Difficult to compare participant scores on CEEA because of variables within and between sessions. Different researchers used, and no inter-rater reliability test. Volunteers found AS observation form used confusing and open to misinterpretation – inconsistency between volunteers. Didn't know what level of detail required.	Key consideration was not to distress or over-burden participants with onerous evaluation requirements – result was that some planned evaluation activities not carried out. CEEA tool lacks sensitivity required, in part because it only takes one measure across an entire intervention.
[8] Franklin-Gould (2013). UK	To explore the impact of a variety of different arts-based group/workshop projects for people with dementia (n=128) and carers (n=81) in London.	Mixed methods. Post-workshop questionnaire for participants. Quality of life assessed using smiley-face scale (confidence, energy, socialising). Monitoring. Questionnaire for facilitators. Observation. Follow-up 3 months post project.	None reported	PWD and carer answered separately but were questioned together. Responses noted by volunteers, facilitators. Questionnaires developed following consultation with researchers and others. No details of analysis methods, except for mention of themes.
[9] Fujiwara & Lawton (2015) UK	To assess the cost effectiveness of a 24-week weekly programme of group reminiscence arts workshops for people with dementia living in residential care settings.	Cost effectiveness analysis, using Green Book guidelines and data from evaluation study (reported in Nicholson et al. 2015).	Randomising selection of control sites not possible due to uncertainties around availability within required timeframe. Ethics committee suggested interventions be distributed among care homes on like for like basis. Estimated trend line may overstate scores in control group because dementia conditions can deteriorate rapidly over longer time spans. Sample attrition. Parallel trends assumption not tested because only one period of data collected before programme started.	CBA study not possible with chosen outcome measures. Measure of relative programme effectiveness that can be compared against results for other dementia related programmes. But currently there are no such estimates using same outcome measures. This is a key area for future research.

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
			Other benefits to programme not captured by chosen outcome measures, and therefore analysis may understate benefits.	
[10] Garabedian (2015) UK	To explore the effects of individualized live and recorded music on people with dementia and their carers when the person with dementia was in the final phase of life.	PhD thesis. Interviews, surveys. Video footage analysed using PIECE-Dem. Researcher conducted intervention.	Manager support needed to ensure staff support. Busy care staff. Ethos of care homes beyond control of researchers. Researcher viewed as 'outsider'. Biological measurements proved intrusive and results were not useful. Staff interruptions were frequent. Single researcher.	
[11] Gant, Hetherington & Reynolds (n.d) UK	To evaluate the impact of a pilot programme of creative art initiatives for people with dementia designed to improve wellbeing and socialization, and taking place in community and care settings. Focuses on value of work for people living with dementia, the need, measures of success, connections made and improvements in wellbeing. Also explores implementation issues.	Case studies. Reflective journals, interviews, observation, participatory evaluation session.	Detailed observation found to be most effective approach while direct questions asked of participants at start and end of session using word cards not found effective. Participants may not have language or memory to describe feelings. Informed consent tricky. Evaluation methods can become intrusive. Beneficial outcomes diverse because individuals have differing needs and interests.	Beyond scope to measure specific health outcomes. Ethical issues include understanding need for confidentiality, permission to participate, and informed consent for those at advanced stages. Recommendations include: involve as many people as possible in assessing impact, and encourage use of common framework (eg Five Ways to Wellbeing). Value of observation. Develop links between practitioners and researchers. Adopt ethical guidelines, recognizing potential

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
				for negative impact. Disseminate findings widely. Recommend: 'practitioners as researcher' role. Carry out informal observations in advance to enable changes to be more effectively captured (benchmarking). Draw on existing models of good practice for health and social care. Recognise value of observations by artists. Flexibility.
12 Gregory & Windle (n.d) UK	To evaluate the impacts of a 10 week intergenerational programme of art sessions for people with dementia and their carers on procedural memory, communication and mood, and attitudes towards people with dementia.	Questionnaires for participants/carers at beginning and end. Focus group. Observation by artists. Follow up interview with artists. Thematic analysis.	Further comparison with similarly implemented socialization group would help demonstrate the benefits of art in particular.	
13 Hara (2013) UK	To explore how 'musicking' can be used locally to support people with dementia and their caregivers.	PhD thesis. Ethnographic. Participant observation. Grounded theory.	Reflexive observations throughout.	Continuous involvement with participants prepared researcher for interviews. Mutual trust built up and helpful in understanding meaning of specific habits and character. Participants sang during interviews and singing could be a way to elicit data and make participants feel comfortable.
14 Harries, Keady & Swarbrick (2013) UK	To evaluate the impact of a multi-art form project incorporating drama, storytelling, singing, poetry and crafts, for people with	Qualitative. Artist reflections and observations. Interviews by artists with participants. Feedback forms pre, mid and	Artist found it difficult to keep recording observations during course of workshop. Too many things to note.	Focus on process and participation in the project, rather than determining 'hard outcomes'. Overly clinical methods risk losing sight of context and importance of

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
	dementia. To understand how creative activity can best be delivered for people with dementia and carers. To explore the effects on quality of life.	post. Creative outputs. Photographs.		flexibility in approaches to dementia care. Don't help understand how creative activity has effect. PWD don't maintain consistent levels of attention and improvement. Change particularly difficult to observe in hospital environment. Many studies don't take account of reported experiences of individual participants. Observation methods on their own raise issues around subjectivity and may lead to misinterpretation. Flexible methods suit flexible implementation of programme.
[15] Hirst, Lane & Mikelyte (2012) UK	To explore and evaluate the impacts of a series of seven weekly performance-based workshops for residents of a dementia care facility	Mixed. Observational data, including post-session. Film. Quant recording of participants' actions and emotional states. Conversations with staff.	Pilot project with small sample size. No control. Observational only. Observations before and after sessions were of such a different kind of activity and context that it is difficult to compare.	Maintenance work disrupted sessions, meaning participants had to move to different room for one session.
[16] Irving (2011) UK	To evaluate whether participation in creative activity can help combat problems of isolation and loneliness, and improve wellbeing of older people with dementia or depression. To assess strengths and weaknesses of management and delivery of project, impact on wellbeing for participants, suitability of chosen outcome	Interviews with stakeholders. Observations. Interviews (n=15) and focus groups (n=5) with participants and carers. Attendance records and artists diaries. LWQ, MHC outcomes assessment tools. Thematic literature review.	Participants have memory recall problems, impaired capacity for judgements, may have impaired sensory, motor and communication skills. Evaluation methods must be sensitive and tailored to needs. Difficulties with LWQ: most participants unable to self-complete, made them feel anxious, and required support; participants struggled to understand concepts and became confused by apparent repetition; overwhelmed by volume of	Hoped that referral GP would be able to undertake analysis of impact of workshops on level of contact with health professionals and medication – but sample size to small and timeframe too short. Effect of significant events in individuals' lives on level of change.

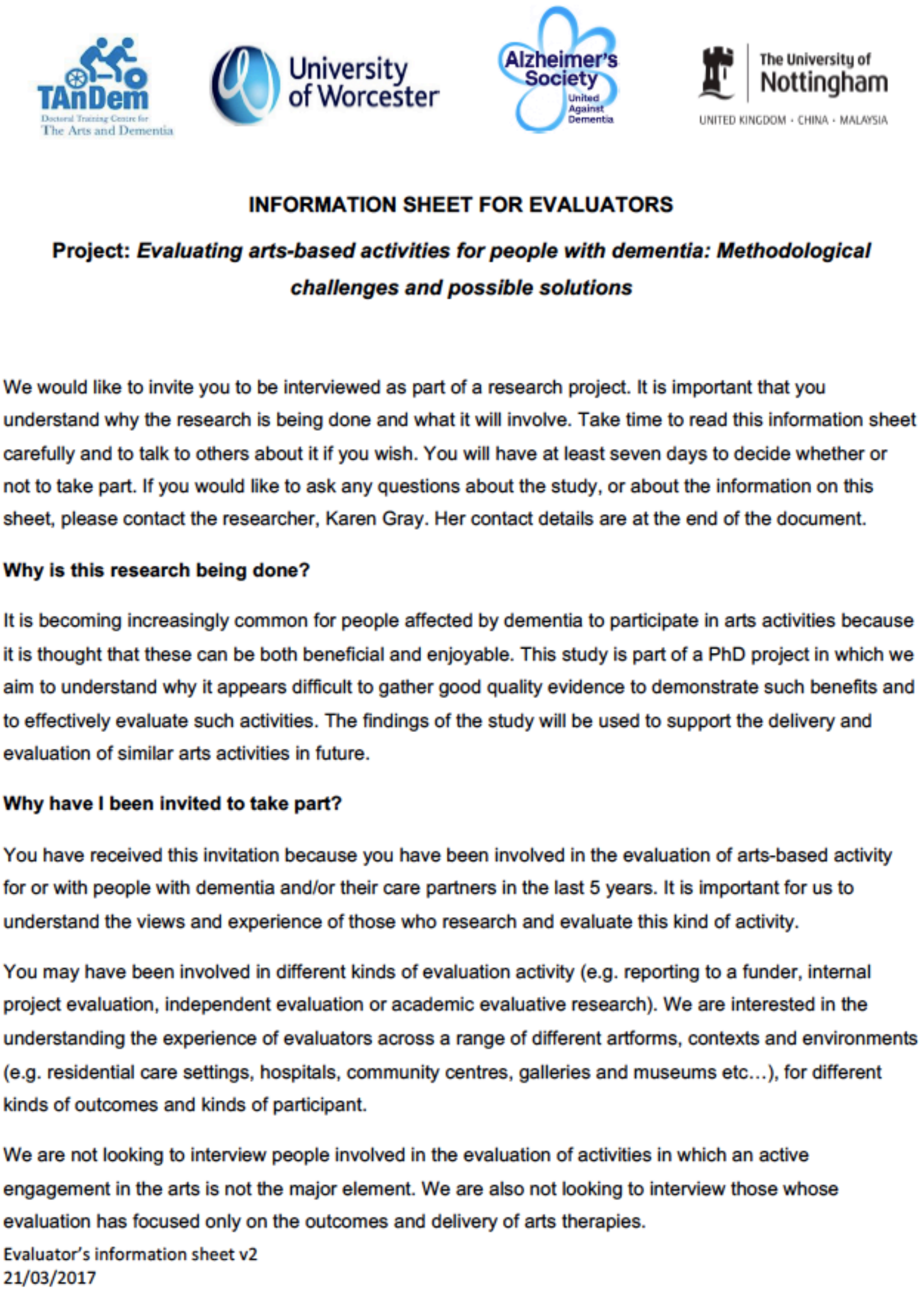
Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
	tool and key learning points for future.		questions and became tired; struggled to read small font, layout and grey colour scheme; some had insufficient motor skills to hold a pen.	Developed 'board game' version of LWQ and trialled with one participants. Participants enjoyed completing MHC tool designed specifically for use with PWD.
[17] Johnson (2015) UK	To investigate the impact of museum object handling and art-viewing in a museum on the subjective wellbeing of people with dementia and their carers.	Quasi-experimental repeated measures crossover design. Pre/post. N=66 PWD and carers.	Sample not representative, included over-representation of people who attend groups, excluding those who do not. Prior interest in art. Gender imbalance: most PWD were male, carers female. Individual differences – participants responded differently, no way of knowing why. VAS was limited in scope, and depended on ability of participants to conceptualise it as representation of abstract concepts. Participants may give higher ratings when VAS are vertical. Short term.	Future RCT planned. Recommends use of observational methods, including video footage, brain-scanning. Broaden scope to include people with severe dementia.
[18] Kimber-McTiffen (2012) UK	To evaluate a programme of weekly dance classes intended to improve the wellbeing of people with dementia attending a specialist day care centre.	Observation, feedback, discussion, photographs from sessions.	Difficulty of 'proving' long-term benefits because of nature of condition. Measurements of frequency of interaction, rather than quality.	No details of analysis or data collection methods. Mixed aims. Key concepts (eg wellbeing) undefined. Mis-use of terms 'evidence' and 'proof'. Thoughtful description of intervention and implementation.
[19] Lloyd (2016) UK	To explore the role of a three year reminiscence arts programme in the lives of care home residents living with dementia.	Practice-based research.	Reflexive observations.	Inspired by relationship-centred approach to care. Struggle to articulate in words some aspects of artwork and research process.
[20] London Centre for Dementia	To evaluate the effects of an interactive theatre residents on	Qualitative. Film, photos, surveys, reflective dairies0,	None reported.	Priority given to observing and talking to residents and staff about

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
Care & Central & Cecil (2010) UK	residents and staff at a care home for people with dementia.	focus groups, interviews, structured observation. Validated measures for wellbeing (Bradford Dementia Group Wellbeing Profiling tool) (n=8) and staff attitudes to dementia. Baseline and 6 months post.		their behaviours, feelings and experiences. Residents assessed selected by staff as representative of current residents.
[21] Minghella (2017) UK	To understand the impact of a pilot dance programme, delivered by professional dancers, working with groups of people with dementia in two care settings. To explore the experience of participating, and the impact on all those involved. Also explored learnings for future implementation.	Formative. Time series. Mixed methods, including non-participant observation, photographs, video and interviews. Bespoke questionnaires assessed QoL.	Other factors may have contributed to changes in scores. Figures based on average and aggregated scores. Individual ratings showed more variation, e.g. individuals rarely improved in all areas.	
[22] Morphew & Murdin (2013) UK	To evaluate a programme of music, storytelling and conversation sessions delivered in the elderly care ward of a hospital, with the artists working with patients with dementia in particular.	Artist reflective journals, with staff survey pre and post. 2 case studies of patients who attended over the longest period of time. N=70 patient participants.	None reported.	Difficult practically to deny access to patients who did not have dementia. Patients had multi/complex needs
[23] Musica (2016). UK	To evaluate the impact on wellbeing, memory and social relationships of a music programme for people with dementia resident in care homes.	Feedback, observations and questionnaires. DCM.	None reported.	
[24] Nicholson et al. (2015)	To evaluate a 24-week weekly programme of group reminiscence arts workshops	Mixed methods. Multidisciplinary. Intervention and wait list, 6	Limited staff and care home manager engagement with evaluation. Care staff found questionnaires difficult to understand, lengthy	Research methods from arts provide context for quant through critical engagement with the 'how'

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
UK	for people with dementia living in residential care settings.	homes in each. Comparative, time-series. N=6 care homes assigned to intervention and n=6 to wait list (treatment as usual). 6-8 residents in groups in each of the 6 care homes. Quant: DCM used as outcome measure, at baseline and every 3rd week, + 3 month post intervention. Staff surveyed to assess wellbeing, stress and burnout at baseline and 24 weeks. Qual, ethnographic and practice-based, to explore embodied experience of participants – supplemented by interviews and focus groups with staff and managers.	and intrusive and avoided completing. Some did not have English as first language. Therefore WES questionnaire not used at follow up. Concerns over confidentiality. Consent issues. Multiple layers of authority within care homes. Some managers not engaged, some care staff did not understand or were not aware of programme or evaluation. Some felt unsupported by managers. Some queried research and appropriateness of resident attending group – particularly in control arm of study. Care staff unsure about diagnosis of dementia. Research team commissioned after intervention finalized. Expectation that evaluation would produce proof that intervention worked. Qual researcher sometimes needed to participate in groups being observed by DCM observer. Research included different combinations of care homes, findings recorded at different intervals. Timescale for ethical approval and urgency of producing results in time to attract further funding was pressure on evaluation.	and 'why'. Quant provides evidence of effects. Findings from different disciplines are presented largely as separate studies. Projects tied into funding cycle, uncertainty about continuity. Care homes closing, financial impact on commissioning. Notes impossibility of completely understanding another person's experience, particularly that of person with dementia.
[25] Tyack (2015) UK	To investigate the impacts of a novel touchscreen art-viewing application on the wellbeing of people with dementia.	Doctor of Clinical Psychology thesis. Mixed methods exploratory study. Measures of wellbeing before and after intervention and before and after each art-viewing session. Interviews. Thematic analysis.	Small sample size, cannot generalize. No control group. Five sessions seems insufficient to determine when increases in wellbeing might level out or diminish. Not all pairs used the tablets at least five times. Not possible to know how the tablets were being used. Participants self-selected, members of group who attend Dementia Cafes. Volunteers, meaning tend to be more optimistic and resilient than average.	
[26] Vella-Burrows (2016)	To assess the effect of two dance programmes (one in community, one in hospital) on	Two cohort, repeated measures, with case studies. Mixed methods. Adapted	Selective sampling. Small sample size. Lack of control for confounding factors. Generalisability not possible.	QOLAD not used in hospital environment because of potential for overburdening participants.

Study	Study aim	Design, methods and tools	Limitations and challenges cited	Notes
UK	the quality of life and wellbeing of people with dementia taking part, and whether it might provide a sustainable model of dance activity that could be facilitated by healthcare staff.	scales for wellbeing, QoLAD, Prosper Involvement and Prosper Wellbeing. Participant observation, video, interviews and focus groups. Study steering group included older people, n=4 affected by dementia, healthcare professionals and researchers. Data tools piloted and revised with group. Thematic analysis. Video analysed using thematic framework template by clinical specialists.	Prosper Involvement and Wellbeing Scales not yet sufficiently refined. Need further development.	Future research should include cost benefit analysis. Recommends larger samples and longer period of time
[27] Waugh (2015) UK	To evaluate the ways in which a pilot programme of arts and reminiscence activity contributed to the increased wellbeing of care home residents with dementia.	NEF wellbeing outcomes framework. Surveys at end of workshops. Observation by evaluator of first and last sessions. Interviews. Monitoring data.	Could not apply NEF Wellbeing durational survey because of time and staff constraints, severe illness of participants, large size of groups, poor continuity of attendance.	Attendance figures fluctuated. Participant groups were not the same each week. Care staff applied survey. Anecdotal observations.
[28] Whitaker (2016). UK	To detail the development, implementation, and activities of a music programme involving professional musicians working in groups with people with dementia living in care.	Feedback. Reflections. 'Case studies' of participants. Staff observations.	None reported.	
[29] Young (2014) UK	To explore the impact of two 8-week programmes of art-viewing and art-making in a contemporary art gallery setting on the cognitive skills of people with dementia.	Doctor of Clinical Psychology thesis. Quantitative. Exploratory. Content analysis of audio recorded and transcribed sessions.	Not always possible to identify individual participants from recordings, therefore participant data grouped, no individual level analysis. Did not address potential differences between groups. Self-selected sample, prior interest in arts. Malfunctioning recording equipment. Potential recency effect – artmaking always followed art-viewing.	Recommends incorporation of usual activity control group, larger sample sizes, measurement of further cognitive skills.

9.6 Example participant information sheet



INFORMATION SHEET FOR EVALUATORS

Project: *Evaluating arts-based activities for people with dementia: Methodological challenges and possible solutions*

We would like to invite you to be interviewed as part of a research project. It is important that you understand why the research is being done and what it will involve. Take time to read this information sheet carefully and to talk to others about it if you wish. You will have at least seven days to decide whether or not to take part. If you would like to ask any questions about the study, or about the information on this sheet, please contact the researcher, Karen Gray. Her contact details are at the end of the document.

Why is this research being done?

It is becoming increasingly common for people affected by dementia to participate in arts activities because it is thought that these can be both beneficial and enjoyable. This study is part of a PhD project in which we aim to understand why it appears difficult to gather good quality evidence to demonstrate such benefits and to effectively evaluate such activities. The findings of the study will be used to support the delivery and evaluation of similar arts activities in future.

Why have I been invited to take part?

You have received this invitation because you have been involved in the evaluation of arts-based activity for or with people with dementia and/or their care partners in the last 5 years. It is important for us to understand the views and experience of those who research and evaluate this kind of activity.

You may have been involved in different kinds of evaluation activity (e.g. reporting to a funder, internal project evaluation, independent evaluation or academic evaluative research). We are interested in the understanding the experience of evaluators across a range of different artforms, contexts and environments (e.g. residential care settings, hospitals, community centres, galleries and museums etc…), for different kinds of outcomes and kinds of participant.

We are not looking to interview people involved in the evaluation of activities in which an active engagement in the arts is not the major element. We are also not looking to interview those whose evaluation has focused only on the outcomes and delivery of arts therapies.

If you take part, you will be one of at least five evaluators involved in the study. Other research participants will include people with dementia and their care partners, arts practitioners and facilitators, and those who fund and commission arts activity for people with dementia.

Do I have to take part?

No. It is up to you, and please take your time to decide. We will wait at least seven days and if we haven't heard from you, we will be in touch just once after this time to ask if you need further information to help you decide. You can decide not to take part at any point before the interview takes place. You can also ask for your information not to be included in the study up to 14 days after the interview. If you decide to take part, you will be asked to sign a consent form. If you wish to withdraw you should contact the researcher at the address given below.

What will happen if I agree to take part?

If you tell us you are interested in taking part, Karen Gray, the PhD researcher on this study, will contact you. She will make sure there are no remaining questions you might have about the study and your involvement in it. She will make sure you have the experience we need for you to be part of this research. She will then need to make a final decision about whether or not we can include you in the study. You will be told this decision before 7 April 2017.

If you do take part in the study, we will then set a date to interview you, either on the phone or in person. This interview will take place at a time or a place that is convenient to you. You should not have to travel for the interview, but if you do, we will pay travel expenses. If the interview takes place on the phone, we will call you. This interview will take place between 31 March 2017 and 19 May 2017. You will need to sign a consent form before the interview takes place.

The interviewer will be Karen Gray. She will audio-record the interview, which will last around 45 minutes. The interview will then be transcribed in full so that it can be used in our research.

It is possible that we may wish to interview you again. If do we will tell you about this later and you will have a chance at that time to decide whether you wish to be interviewed again or not.

What kind of questions will I be asked?

You will be asked about your experience of evaluating the delivery and outcomes of arts activity with and for people with dementia. You will be asked your views about evaluation in this context.

Will the information I give stay confidential?

Yes. We will collect two different kinds of information from you.

(1) Your name and contact details. These will be kept securely at the University of Worcester. In line with Data Protection principles, they will be kept for only as long they are required after the completion of the PhD project (expected to be December 2018). Only the researcher, Karen Gray, will have access to this information, which will be used only to contact you during the study.

(2) An audio-recording and transcript of the interview with you. These will be stored securely as digital files and will be used only for research purposes by the research team. We will also keep a descriptive note of your role and practice and where you generally work (e.g. 'independent evaluator primarily involved in evaluating music and singing activities delivered in residential care settings'). This information will be linked to your interview data but both will be anonymised. With your permission, we would like to keep this information indefinitely so that it can be used for further research. Your time, expertise and experience is valuable and we would like to make sure that it can continue to be valuable in future. However, your information will only be used with the permission of the research team members, and only if its use is directly related to similar research. If you do not wish this to happen, you will have a chance to ask for your information to be securely destroyed 12 months after completion of the PhD project.

If we quote from the interview with you in reports, documents or conference presentations produced as a result of this study, we will not mention you by name or in relation to any named organisation that employs or has employed you, or to the name of any of the settings or projects you have evaluated.

What will happen to the results of the research?

This research is being carried out as part of a PhD project at the University of Worcester. Findings will be published in a PhD thesis written by the researcher, Karen Gray. They may also be published in journals and other publications and presented at conferences and workshops.

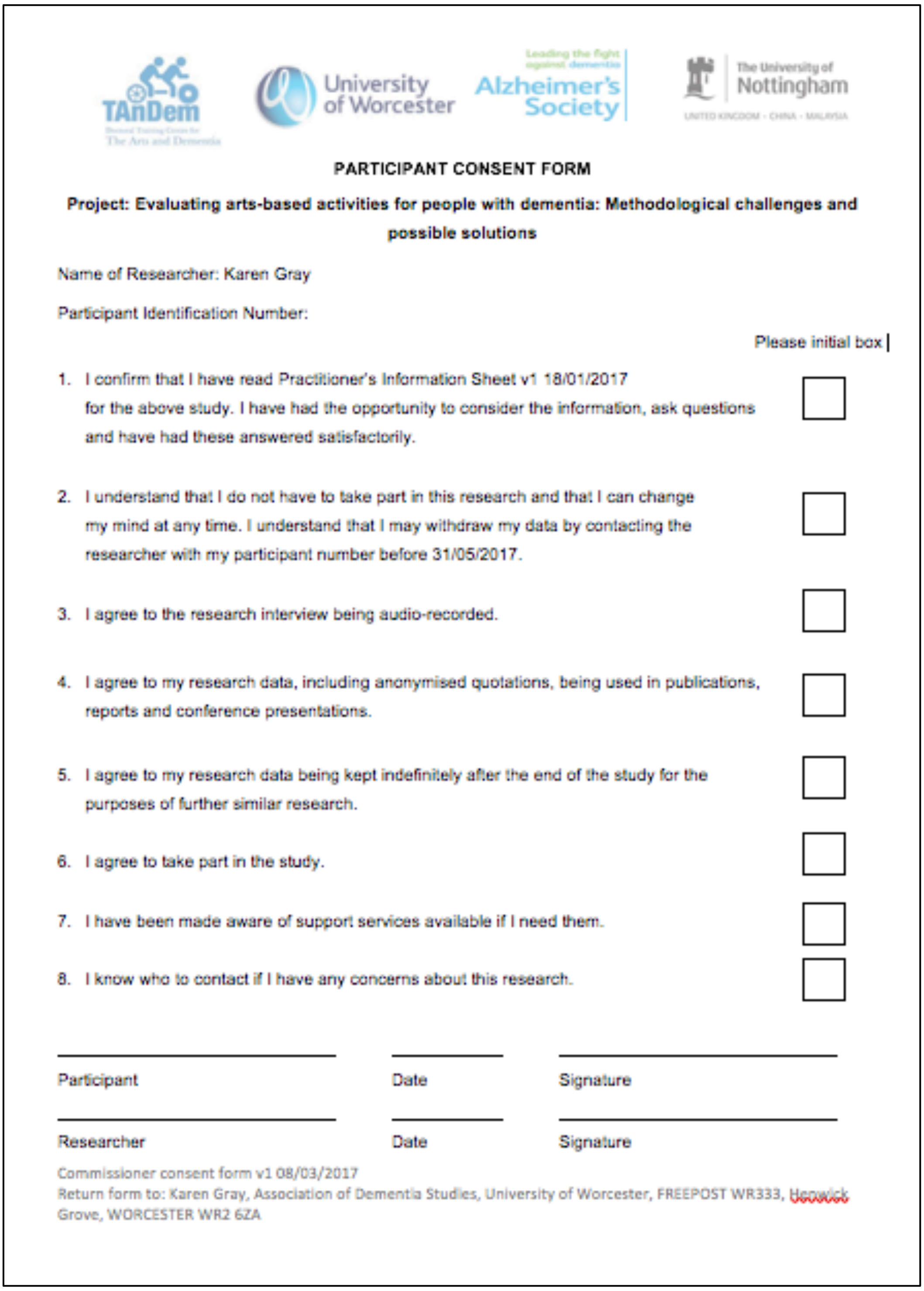
You will have an opportunity to ask us to tell you about outputs of the study, such as those described above.

This project is connected to a larger programme of work being carried out by TAnDem (The Arts and Dementia) Doctoral Training Centre. This is a partnership between the Association for Dementia Studies at the University of Worcester and the Centre for Dementia at the University of Nottingham. It is funded by the Alzheimer's Society. You can find out more at: http://www.worcester.ac.uk/discover/tandem-phd-studentship.html.

What are the benefits and disadvantages of participating in the study?

In taking part in this study you will be supporting research whose aim is to aid the delivery of more effective and more beneficial arts activities for people with dementia. Your expertise and experience are a vital part of this research.

9.7 Example participant consent form



TAnDem The Arts and Dementia
University of Worcester
Leading the fight against dementia Alzheimer's Society
The University of Nottingham UNITED KINGDOM · CHINA · MALAYSIA

PARTICIPANT CONSENT FORM

Project: Evaluating arts-based activities for people with dementia: Methodological challenges and possible solutions

Name of Researcher: Karen Gray

Participant Identification Number:

Please initial box

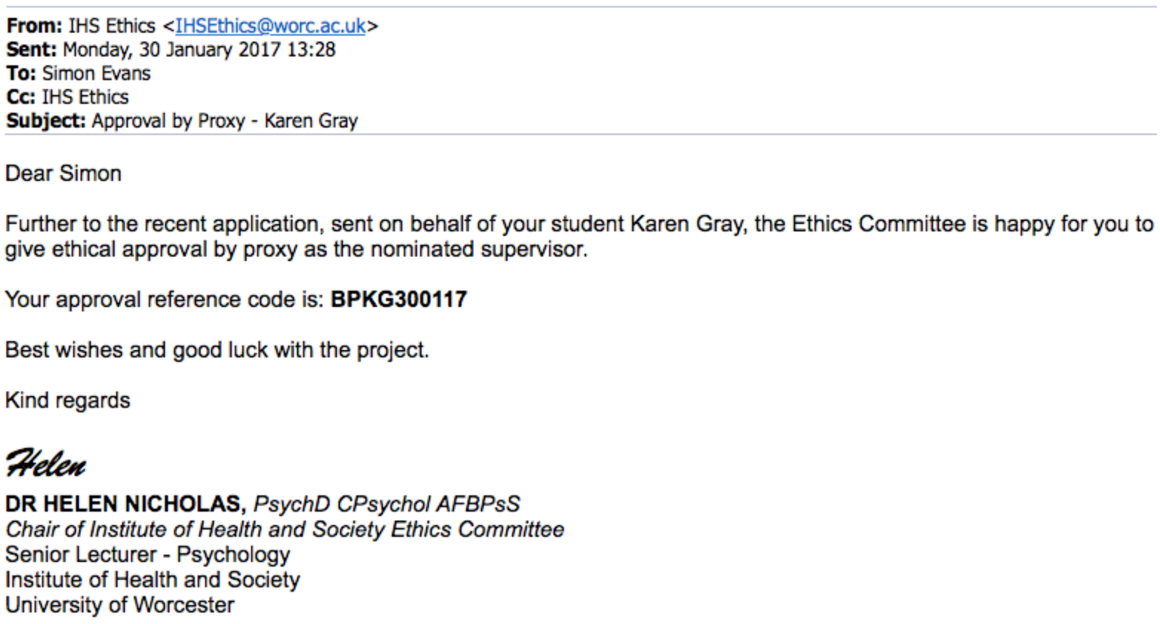
1. I confirm that I have read Practitioner's Information Sheet v1 18/01/2017 for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐
2. I understand that I do not have to take part in this research and that I can change my mind at any time. I understand that I may withdraw my data by contacting the researcher with my participant number before 31/05/2017. ☐
3. I agree to the research interview being audio-recorded. ☐
4. I agree to my research data, including anonymised quotations, being used in publications, reports and conference presentations. ☐
5. I agree to my research data being kept indefinitely after the end of the study for the purposes of further similar research. ☐
6. I agree to take part in the study. ☐
7. I have been made aware of support services available if I need them. ☐
8. I know who to contact if I have any concerns about this research. ☐

Participant ________________ Date ________ Signature ________________

Researcher ________________ Date ________ Signature ________________

Commissioner consent form v1 08/03/2017
Return form to: Karen Gray, Association of Dementia Studies, University of Worcester, FREEPOST WR333, Henwick Grove, WORCESTER WR2 6ZA

9.8 Confirmation of ethical approval



From: IHS Ethics <IHSEthics@worc.ac.uk>
Sent: Monday, 30 January 2017 13:28
To: Simon Evans
Cc: IHS Ethics
Subject: Approval by Proxy - Karen Gray

Dear Simon

Further to the recent application, sent on behalf of your student Karen Gray, the Ethics Committee is happy for you to give ethical approval by proxy as the nominated supervisor.

Your approval reference code is: **BPKG300117**

Best wishes and good luck with the project.

Kind regards

Helen

DR HELEN NICHOLAS, *PsychD CPsychol AFBPsS*
Chair of Institute of Health and Society Ethics Committee
Senior Lecturer - Psychology
Institute of Health and Society
University of Worcester

9.9 Interview topic guides

Topic guides were specific to the role of the interview in Phase 1, so different guides were produced for evaluators, arts practitioners and funders and commissioners.

9.9.1 Phase 1 (Evaluators)

INTERVIEW TOPIC GUIDE (EVALUATORS) v21/03/2017

Project: Evaluating arts-based activities for people with dementia: Methodological challenges and possible solutions

Researcher: Karen Gray

These are guidelines for structuring the interview. Sub-questions indicate potential prompts. The interviewer will be flexible and allow for other areas of discussion to emerge.

1. Can you give me an overview of your involvement in the evaluation of arts activities delivered for or with people affected by dementia?

    - Some people have said that it is particularly challenging to evaluate arts activities of this kind. How do you feel about this?

2. How would you define ‘evaluation’?

3. Now, thinking about a particular project or activity and its evaluation, can you tell me about this evaluation and your role in it? [Name the project, if known or agreed beforehand]
    - What were the aims of the evaluation?
    - And how was it carried out?
    - What influenced the decisions that were made about these things?
    - Were there any difficulties that had to be overcome?
    - Is there anything about it that you would do differently now?
    - Did the fact that you were assessing an arts activity affect your evaluation?
    - Did the dementia context affect the evaluation?

4. Are there any evaluation methods or approaches that you would like to use in future similar projects?
    - Or are you aware of any that you think might be fruitful for evaluators to explore?

5. Is there anything else that you would like to talk about?

9.9.2 Phase 1 (Arts practitioners)

INTERVIEW TOPIC GUIDE (ARTS PRACTITIONERS) 21/03/2017

Project: Evaluating arts-based activities for people with dementia: Methodological challenges and possible solutions

Researcher: Karen Gray

These questions provide guidelines for structuring the interview. Sub-questions will function as potential probes. The interviewer will be flexible and allow for other areas of discussion to emerge.

1. Can you tell me about the work you do for or with people affected by dementia? [differentiate between an organisation managing arts practice and arts practitioners]
    a. What is your ‘practice’?
    b. What kind of settings and contexts have you worked in?
    c. What kind of participants have you worked with?

2. I'm interested to know more about how this kind of work is evaluated and your experience of evaluation. Thinking generally:
    a. What does 'evaluation' mean for you or for your organisation?
    b. If you are an arts practitioner, do you do any kind of evaluation or reflective practice for yourself?
    c. If you represent an arts organisation, do you require artists working with you to do any kind of evaluation?
    d. Some people have said that it is challenging to evaluate arts activities of this kind. What do you feel about that?

3. I'd like to talk in a little more detail about a particular project or activity and its evaluation [name it if known already, otherwise prompt for a particular project]:
    a. What, if anything, was your role in this evaluation?
    b. Can you describe how this evaluation affected you as a practitioner or facilitator delivering the activities, if it did?
    c. Is there anything that you think could have been done differently?
    d. What do you know about why the evaluation or research was being carried out?
    e. Do you know what, if any, conclusions it reached?
    f. Were these conclusions valuable to you? How?
    g. Do you feel that the evaluation was valuable or valued by others?

4. Is there anything else that you'd like to tell me about?

9.9.3 Phase 1 (Funders / commissioners)

INTERVIEW TOPIC GUIDE (FUNDERS AND COMMISSIONERS)
21/03/2017

Project: Evaluating arts-based activities for people with dementia: Methodological challenges and possible solutions

Researcher: Karen Gray

These are guidelines for structuring the interview. Sub-questions will function as prompts. The interviewer will be flexible and allow for other areas of discussion to emerge.

1. **Can you tell me a little** about your involvement in funding arts and dementia activity**?**
    a. Why do you fund or commission this type of work?
    b. What might influence your decision to fund or commission a particular arts activity?

2. I'm interested to know more about evaluation that is carried out for activities that you fund or commission.
    a. For example, do you have any particular requirements of evaluation?
    b. And, if so, what influences you in setting these requirements?
    c. Does it make any difference that it is an 'arts' activity that is being assessed?
    d. Does the dementia context affect the evaluation?
    e. Some people have said that it might be particularly challenging to evaluate arts activities of this kind. What do you feel about that?

3. Is there anything else that you would like to talk about?

9.9.4 Phase 2 interviews

Key: A&H (Arts and health specialists); A&D (Arts and dementia specialists); D (Dementia researchers)

Topic guide for interviews with evaluators / researchers with experience across the arts and health field / dementia specific		
Question	Logic	Type
I am talking to you because you have experience of evaluation and research from across the arts and health / dementia research field, but some questions will ask you to think about arts and dementia practice in particular. If you don't feel you can comment on this specifically, please relate your answer to arts and health / dementia research more generally.		
Could you briefly summarise your background and research interests in relation to dementia? Have you ever been involved in researching or evaluating any arts and dementia practice? • If so, can you tell me a little about that? • Do you have any thoughts about how it fits into the wider dementia 'picture'? • What do you see as the value of arts practice for people affected by dementia? • What characteristics of arts and dementia practise might you expect to create problems for evaluators? Something I'm interested in exploring is what people perceive as the difference between evaluation and research. • What methodological challenges might you associate with evaluation in particular?	Introductory questions (select as appropriate)	ALL
Some people involved in evaluating arts and dementia practice have said to me about this kind of evaluation: 'that's just the way it works'. They might be talking about things that seem obvious to me about how we might go about evaluating, the funding and resourcing of the work, for example, or the requirement for evaluation to 'speak' to particular audiences. If someone said, about evaluation and research in the field of arts and health / arts and dementia / dementia 'that's just the way it works' – what kind of factors, structures or influences, might they be referring to? What do you think might be the effect of this way of thinking, or these factors on the way in which evaluation and research is conducted?	Exploring characteristics of the infrastructure around arts and dementia - commissioning background (Context)	ALL

Topic guide for interviews with evaluators / researchers with experience across the arts and health field / dementia specific		
There seems to be a real diversity of practice currently being delivered in arts and health work. - What is the impact of this for evaluation?	Exploring effect of diversity of practice (Context)	A&H or A&D
It seems that there is what I will call a 'spectrum of evaluation' for arts and dementia practice. By this I mean that there is some academic evaluation research into particular programmes, evaluation is conducted on a consultancy basis by academic or other independent evaluators, it happens internally within organisations in different ways, and it might also be engaged in by arts practitioners themselves as part of their own reflective practice). - Would you agree with this characterisation? - If so, what, if any, do you feel might be the impact of this for evaluation?	Exploring spectrum of evaluation practice (Context)	A&H or A&D
Particular kinds of outcomes are often associated with arts and dementia practice. In particular, there is a lot of emphasis on wellbeing, quality of life and the impact on behavioural and psychological symptoms of dementia - What challenges might you associate with work that focuses on these kinds of outcomes? People have commented that arts and dementia practice involves effects that are 'intangible': - What do you think lies behind this statement? - What effect do you think this might have on the way in which evaluation is conducted? Do you think that arts and dementia practice is currently being driven by any particular conceptions of what is valuable? - What effect do you think this might have on the way in which evaluation is conducted?	Exploring anticipated outcomes (Context) Exploring attitudes towards value and sense that effects are intangible (Mechanisms)	A&H or A&D (& DEM if appropriate)
Arts and dementia practice is often delivered in quite specific settings. For example, health and care settings such as hospitals or residential care homes. - How do you think these settings affect its evaluation?	Exploring settings (Context)	ALL
How do you think the condition of dementia itself might prove challenging for evaluators? I am thinking about both the nature of the condition, but also the way in which people affected by dementia are treated or cared for.	Exploring effects of condition of dementia (Mechanism)	A&D and D
Evaluators I have spoken to talk about experiencing a number of ethical difficulties. - What difficulties would you expect them to experience?	Exploring ethics (Mechanisms and Outcomes)	A&D and D

Topic guide for interviews with evaluators / researchers with experience across the arts and health field / dementia specific		
- It has been suggested that the formal ethical approval process is not suited to the kind of contexts and situations in which evaluators of arts and dementia practice find themselves working. I wonder if you have any thoughts about this? - What might be the effect on evaluations of the kind of difficulties we have been talking about?		
I am exploring the idea that some of the challenges for evaluators stem from the way in which different 'world views' relate to each other. For example, it been suggested that arts and dementia practice is experienced by those directly affected by the condition, is often lead or devised by the arts world, evidence about its effects is required by or directed at the health or dementia care world, but may be produced by evaluators from yet another background. - I wonder if you have any thoughts about this? - Further to these thoughts, evaluators themselves may also be from different 'disciplines'. What kinds of effects might you expect to see resulting from these differences? - Do you think there are ways in which the different groups represented might work together to evaluate the work more effectively?	Exploring effects of different world views / disciplines (Mechanisms). Exploring potential solutions (Outcomes)	A&H or A&D
Artist and arts organisations often say about arts and dementia practice – 'we all know it works'. - What do you think is behind this statement or attitude? - What effect do you think it might have on evaluation?	Exploring lack of understanding about how art works, lack of theorisation (Mechanisms and Outcomes)	A&D
When arts practitioners talk about the impact of their work on people living with dementia, they often do so with stories. - I wonder if you have any thoughts about why this might be the case? - How appropriate do you feel it is to talk about 'story' when evaluating arts and dementia practice? - Do you think there are ways in which story might be appropriately used when evaluating arts and dementia practice? - Do stories have a place in the way in which evaluation takes place in the dementia field?	Exploring stories, nature of evidence, advocacy. Potential solutions. (Mechanisms and Outcomes)	A&H or A&D D
Artists often reflect on their own practice in particular ways. One of the things I am exploring is how this reflection or its process is or might be used in evaluation. - I wonder if you have any thoughts about this?	Exploring the use of reflective practice in evaluation – potential solutions	A&H

Topic guide for interviews with evaluators / researchers with experience across the arts and health field / dementia specific		
- Do you think there are ways in which reflective practices might be appropriately used when evaluating arts and dementia activity?		
Are there any methodological practices or developments for evaluation of arts and health practice / dementia research that you see as particularly important at present? - Are there any that seem particularly pertinent for arts and dementia practice?	Wrapping up question	ALL

9.10 Poster: Arts and dementia evaluation: addressing some of the challenges

ARTS and DEMENTIA EVALUATION:
addressing some of the challenges

Based on original research by Karen Gray, this is a graphic narrative created by ForMed Films. It is based on interviews with 26 artists, evaluators, managers, funders and commissioners of arts activities for people living with dementia. Interviews were conducted as part of a study exploring the methodological challenges of evaluating these activities, with the aim of identifying some solutions. While the words are all theirs, the experiences are not those of just one individual. This research was funded by the Alzheimer's Society as part of a PhD studentship within the TAnDem Doctoral Training Centre.

I'm looking for outcomes like:

- INCREASED INTERACTION
- RELAXED
- INCREASED WELLBEING
- FEELING GOOD
- ENHANCED MOOD
- SELF EXPRESSION
- ABILITY TO COMMUNICATE
- REDUCED ANXIETY

FUNDER

The arts, by a lot of people in social or health care, is still not perceived as something more meaningful, in general.

It is crucial that arts organisations do in depth research and evaluation of the work that we're doing because all of those outcomes are going to be lost.

I'm looking for outcomes, But also to integrate it into the practice of the home. Getting the good information, and the reports, into care plans.

ARTIST

Our aim is for a person with dementia to have an experience of themselves as creative, as connected,

so much of it is intangible isn't it?

There are some moments where something special happens, and it's really difficult to fit that into some of the more conventional ways of evaluating the work.

EVALUATOR

It's a very pragmatic area. The projects aren't designed with evaluation in mind,

There's a desire to have a long-term change and impact. Having a good time this morning is one thing but does that impact people's behaviour longer term?

Is that the right question to be asking? because the here and now is so important.

If you think you're going to get some tidy little charts from this... your going to be very disappointed!

We need a much bigger sample to verify.

You can't just EVALUATE for the sake of it!

ASK QUESTIONS

An organisation has to decide WHY it wants to do the evaluation, what purpose is it for?

FIND OUT THE VALUE

A real all-round look at the benefits from the different points of view.

GET THE PARTNERSHIP RIGHT

Create an atmosphere where there's good communication and equality throughout.

Different disciplines learning from each other.

A COMMUNITY OF PRACTICE

It's about the relationship that people with dementia have with each other, and also with the care staff, and also with us

It's about CONNECTIVITY!

BE REFLECTIVE

How did that go?

Evaluation is the kind of sifting through and seeing what things of meaning have come out that we can learn from and that are going to change the practice.

Try and really SEE THE SUCCESS OF THE PROJECT in the context in which it's delivered.

it was an absolute joy, one of the best research intiatives, absolutely food for the soul!

DON'T JUST IMITATE

STANDARD FORMS

VALIDATED SCALES

DIFFERENT KINDS OF EVIDENCE

Try and find more effective and less time-consuming methods for evaluation.

MAKE MEANING

I END UP TALKING IN STORIES

You need lots of different sorts of evidence. We absolutely believe in showcasing, and heart as well as head.

Research by Karen Gray (University of Worcester). Funded by The Alzheimer's Society. Construction and Illustration by ForMed Films. For more information, contact karen.gray@worc.ac.uk.