

Social
Research
Association

Issue 2
Summer 2016

Social Research Practice

The SRA journal for methods in applied social research



Social Research Practice

Issue 2 Summer 2016

The Social Research Association journal for methods in applied social research

Contents

- 03** **Listening to young children in practice: using and adapting the Mosaic approach**
Cathy Street, Emma Wallace and Puja Joshi, NCB Research Unit and Lucy Williams,
NCB Early Childhood Unit
-
- 11** **Preparing transition from child and adolescent mental health services (CAMHS):
evaluating a creative participatory approach**
Caroline Lee, Cambridge Institute of Public Health, University of Cambridge
-
- 21** **Addressing quality in social media research: the question of representivity**
Josh Keith, Ipsos MORI, Steven Ginnis, Ipsos MORI, Carl Miller, Centre for the Analysis
of Social Media, Demos
-
- 31** **Comparison of clustering effects for primary sampling units**
Kevin Pickering, Patten Smith and Ben Hornsby, Ipsos MORI
-

Editorial

Richard Bartholomew

Editor

*Welcome to **Social Research Practice**, the SRA's journal for methods in applied social research.*

We received a very positive response to our first issue; this second issue builds on that. In this issue, we show the SRA's strengths in representing both qualitative and quantitative research methods.

The first two articles demonstrate how qualitative methods can be used to engage and learn from the perspectives of very young children and teenagers.

In **Listening to children**, Cathy Street and colleagues show how the Mosaic approach to research with children and their parents can be successfully adapted even when time and other constraints make it impossible to use the full model. Some believe that trying to consult directly with very young children is too difficult, and that interviewing their parents is sufficient. The authors show how direct work with very young children can provide valuable insights in addition to those offered by parents.

In **Preparing transition from child and adolescent mental health services (CAMHS)**, Caroline Lee demonstrates how using creative and participatory approaches with young people can help to assess and improve the difficult transition which 17 to 18-year-olds face when they transfer from specialised child and adolescent services to adult mental health services or primary care. This research was closely linked with practice, and two NHS Trusts are now using the findings to improve how they support young people.

The ubiquity of social media (six in ten people now use it regularly and this will grow) presents new opportunities for social research, especially because of its immediacy and low cost. But social scientists have been rightly cautious and sceptical about the nature of the data it generates, particularly its representativeness. Does that mean we should try to keep it at arm's length or, by understanding more about its sources of bias, are there ways to use it for objective research? In **Addressing quality in social media research: the question of representivity**, Josh Keith and colleagues describe their research comparing views expressed on and offline, as well as the extent to which the opinions expressed online reflect the population of social media users.

The final article, **Comparison of clustering effects for primary sampling units** by Kevin Pickering and colleagues, also considers representativeness, but does so in the context of more traditional concerns about the effects of clustering on the representativeness of random probability samples. Using the opportunity provided by the large-scale General Practice Patient Survey, the team has tested the relative merits of using either merged postcode sectors (PCSs), middle layer super output areas (MSOAs) or census area statistics (CAS) wards as the primary sampling units. The results have important implications for those designing or commissioning future large-scale surveys in which a high degree of precision and efficiency is essential.

I would like to thank all members of our editorial board for their support in producing this issue, and particularly those who have refereed articles. My thanks also to our authors for their hard work and insights.

Social Research Practice provides a twice-yearly forum for developing and discussing the quantitative and qualitative methods which SRA members and others use to design, conduct and commission research on social policy and practice issues. We aim to encourage methodological development by helping practitioners to share their knowledge.

We are keen to establish a regular flow of new articles. We publish issue three in December 2016, with a copy deadline for draft articles of 9 September. If you are interested in offering an article, details, guidelines and templates for submissions are on our website. If you have an idea for an article but are not sure if will be suitable, just drop me a line: rabartholomew@btinternet.com.

Listening to young children in practice: using and adapting the Mosaic approach

Cathy Street, Emma Wallace and Puja Joshi, NCB Research Unit and Lucy Williams, NCB Early Childhood Unit

Abstract

It is relatively rare for children under five to be consulted directly about services which may affect them. However, specialist techniques have been developed for conducting listening work with children under five, often in the context of individual service reviews, and these include the Mosaic approach (Clark and Moss, 2011). In 2014, the Office of the Children's Commissioner (OCC) and National Children's Bureau (NCB) sought to incorporate listening to children approaches within research into the role of various services in minimising the impact of low income on children under five and their families. This paper considers the value and potential of listening to children in research. It also considers the implications of adapting the Mosaic approach to respond to time and budget restraints, and what can be achieved with different levels of resources.

Acknowledgement

The research study on which this paper is based was commissioned by the Office of the Children's Commissioner (OCC), with Dr Alison Clark, Dr Cathy Hamer and Judy Miller, NCB Early Childhood Unit (ECU) Associates as advisers.

The full study report: Joshi, P., Wallace, E. and Williams, L. (2015) Young children's and families' experiences of services aimed at reducing the impact of low-income: participation work with children and families is at www.ncb.org.uk/low-income.

Introduction

Background: listening to children in research and practice

Historically, the views of children under five are rarely explored by those considering the policies and services which affect them. However, over the last two decades, the capacity of young children to communicate effectively about their experiences has been increasingly recognised (Paley, 1997). Several methodological approaches, which can be used to ascertain children's views, have been developed including the Mosaic approach pioneered by Alison Clarke and Peter Moss (Clarke and Moss, 2011).

The Mosaic approach

The Mosaic approach has been developed as a best practice method for listening to children in an ethical and high-quality way which recognises children as 'experts in their own lives' (Browne et al, 2013). Mosaic is most commonly used in reviewing individual early years services and has two stages. The first involves children and adults gathering documentation. The second is piecing together information for dialogue, reflection and interpretation. The approach uses observation, child conferencing and using cameras, tours and mapping. The rich combination of data forms a living picture of children's experiences.

In this project, researchers built up the picture from parent discussion groups; in-depth parent interviews; shared information and observation sessions; and child-led activities. These included map-making; tours of the child's neighbourhood or home alongside family members; a diary activity; child interviews; and play-based activities, including props such as a soft toy called 'Monkey', to help the children express their opinions.

Background to the project

There is an emphasis in policy and practice on listening to the views of young children. Section 3.5 of the Childcare Act, 2006 places a duty on English local authorities to consider information on young children's views when planning and providing services. The profile raising, training and support work offered by the NCB-led Young Children's Voices Network (YCVN) 2006-2010¹ also supported many settings to develop listening work, and helped to emphasise young children's participation within early years practice.

Reflecting the increasing number of families with young children in poverty (Field, 2010) and the importance of the early years to children's long-term outcomes (Tickell, 2011 and Clark and Moss, 2011) in 2014, the Office of the Children's Commissioner (www.childrenscommissioner.gov.uk) conducted a programme of research to examine the role of services in reducing the impact of low income on children in the early years. One part of this was a research and consultation project with families of children under five and, reflecting the commissioner's commitment to children's participation, a key priority of this research was to fully include the voices of young children.

The NCB Research Centre and NCB Early Childhood Unit were commissioned to conduct the work, with the research seeking to:

- Explore the impact of living with low income on children and families' lives
- Better understand children and families' experiences of the services they receive or would like to receive, in particular early years, health and housing
- Hear directly from children and families about what works in reducing the effects of living with low income and how services could be improved, and to use this feedback to inform recommendations to those commissioning services

In the research, NCB sought to apply the Mosaic approach to examine the role of a wide range of services, rather than simply individual early years services. The timescale (less than nine months) and range of services, required pragmatic adaptations in applying only certain aspects of the approach during one rather than multiple visits. This article reflects on the value and potential role of children's voices to inform service development, and considers the implications of implementing a more limited Mosaic-informed approach.

Study methodology

Building on main principles of the Mosaic approach, the following considerations informed the methods adopted for the project:

1. Research content should be appropriate to the child participants, for example, taking account of their cognitive abilities and interests. At age three to four, children are able to conceptualise and communicate their immediate experiences but have not developed the capacity to hypothesise in an abstract way about alternative circumstances (Early Education, 2012). For this reason, the research element with children focused on understanding the nature of children's immediate experiences of services. It then drew on parents to comment on gaps in services and hypothesise about the current and potential role of services in mitigating the effects of low income
2. From a research ethics perspective, young children need to be in an environment and with people they are comfortable and familiar with. So, the research took place in early years settings familiar to children, and with a familiar parent/carer and/or practitioner present

¹ www.ncb.org.uk/areas-of-activity/early-childhood/networks/young-childrens-voices-network

3. Fieldwork methods need to reflect how children under five can and prefer to express themselves. For this reason, the project used various strategies to talk with children including props such as soft toys
4. Parents and practitioners (who know the child well) assisting with interpretation is vital. This is because the way young children communicate may be more about their immediate situation and lack the background context which also needs to be understood

Reflecting the above, we adopted the following methodology:

- Four family participation events lasting one day with 15 children and 16 parents at four early years settings. Each day included different activities, with different children participating at different times. The purpose was to facilitate children's feedback about services in a familiar and comfortable setting, and feedback from their parents
- In-home parent interviews with an additional sample of nine parents. The purpose of these was to widen the sample to include families which do not use childcare, and families in different circumstances. (This element did not involve work with children and is not discussed further in this paper.)

Children's experiences of different services were explored in the following ways²:

- Early years setting: group work with two to four children at a time involving child-led tours of the setting, children taking photos and/or drawing things which matter to them, discussion, observation, and contribution from parents and early years practitioners
- Health services: one-to-one sessions between a researcher and a child with the parent present, involving play-based participation activities such as role-play with toys, doctors' kits, story books, observation and discussion
- Housing and the home: one-to-one sessions between a researcher and a child with the parent present, involving activities such as drawing, playing with a toy house, observation and discussion

An early years practitioner with extensive practical experience in working with and engaging with children led these specialist fieldwork activities. This is essential for quality and ethical listening work with young children (National Children's Bureau, 2011).

The views and experiences of parents were explored during the events in group discussions and interviews. These focused on:

- Exploring factors they considered important for ensuring their child has a happy, healthy and safe upbringing and any issues and challenges
- Types of services used in each area
- How well children's needs for play and learning, housing and health are met and suggestions for improvements

Analysis of children's experience

Analysis involved detailed fieldworker debriefings at different points during the study. Feedback about children's views and experiences, identified from the listening work with children, and reported by parents in discussion groups and interviews, was collated for analysis. This used an analysis framework for each of the three service areas, coded into the main themes arising from debriefing discussions.

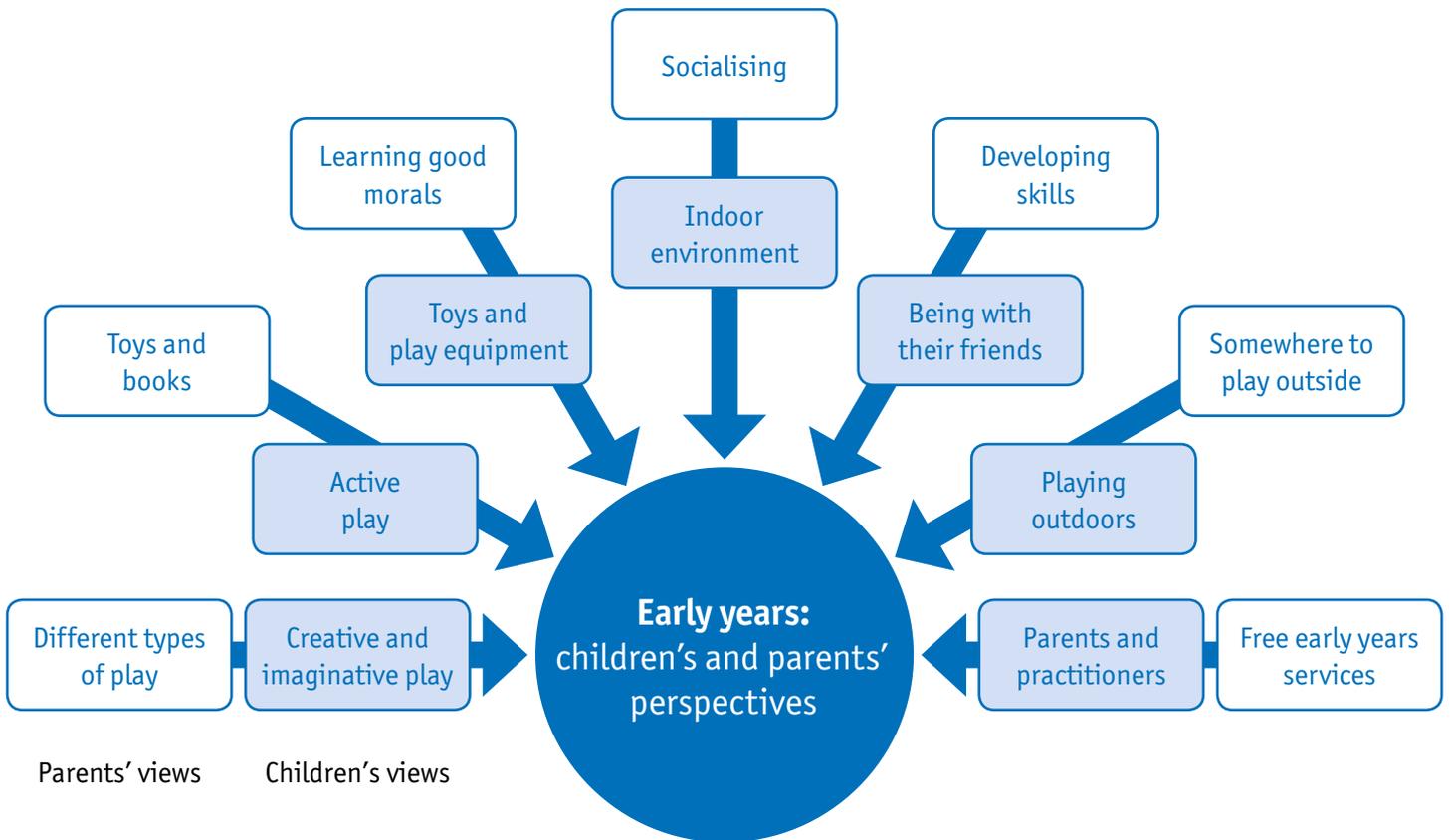
² All activities with children included time for discussion and observation, and subsequent conversations with parents and practitioners, to add context to the children's feedback. Researchers used a soft toy ('Monkey') to help create rapport with children and engage them in activities. For example, during the activities about housing and the home, researchers used Monkey to prompt children to think about the question 'Which rooms would you take Monkey into if he visited you at home?'

Research outcomes from the listening work with children

Understanding children’s needs

When we explored parents’ and children’s perspectives on the factors which are important to children in housing, early years and health, the topics were similar but the respective emphasis of children and their parents was different. Children focused on their immediate experiences, whilst parents were more analytical in discussing the broader role of these things for children’s longer-term development and wellbeing. This is illustrated in Figure 1 below with the inner boxes showing the factors which children identified, and the outer boxes those which parents identified.

Figure 1: Learning and development factors important to children and parents



Children in this study gave detailed information about their lives. This data showed which aspects of services determine how children feel about the quality of what is provided. For housing, health and early years, the most important aspects were people, places, objects and activities, especially those relating to play. This complemented what parents said. Parents discussed children’s needs and the role of services in contributing to children’s wellbeing and development.

Bearing these different groupings in mind may provide a useful indication of what is important for determining children’s comfort and happiness in many situations and service contexts. The summary of feedback from children about their housing in Box 1 illustrates how listening work illustrates children’s world-view.

Children's feedback on housing

Children spoke about their immediate experiences of the home; what they like to do in the home; where and with whom (for example where they play and keep their toys; when and with whom they sleep; how they like playing outside).

Children mentioned rooms or spaces in the home, which were important to them, for example the kitchen, living room and garden. While engaging in a drawing task with the researcher, some children drew themselves in their bedrooms. One parent said that her child's favourite room was the kitchen. Some children mentioned playing outside (for example in the garden, riding their bikes).

Close family appeared important. A few children mentioned playing with their siblings, and others drew their family.

Activities, especially playing, appeared to be a main feature of children's experience of the home. Children discussed some of the things they did at home, for example, 'playing dragons', painting, and playing with dolls. Some also mentioned eating and sleeping.

Contributing to policy recommendations

Discussions with parents highlighted the importance of children having outdoor space; a cross-cutting issue relevant to housing, early years development and health outcomes. Children highly valued outdoor play and they referred to this in conversations about health, housing and early years. Their emphasis on outdoor play corroborated parents' views. This led the study to recommend more join-up between health, housing and early years services to improve children's access to outdoor space and support their well-being.

Engaging with children at a suitable level

The study findings highlighted the importance of taking children's comprehension levels into account when asking them about services or their experiences. For example, for health, children's awareness extended as far as GPs and dentists, and to a slightly lesser extent, hospitals. They knew of these because of personal experience, or from books, television and family members.

Many children understood what happened at health services, and why. Some children mimicked health professionals while using the play equipment, for example, saying 'open wide!' when using the thermometer. One child referred to the 'juice' in the syringe which a doctor would give to their patient. When asked how they would feel after a visit to the doctors, most children said 'better', although many thought that seeing the doctor could be unpleasant. (For example, some children said that the toy injection would hurt, causing an 'ouch!'). This demonstrates that children appeared to have a basic understanding that healthcare is positive, even if sometimes unpleasant. Some of the parents said they were surprised that their children appeared to know more about healthcare than they thought.

These findings highlight that children understand what is happening when they attend health services, and the importance of communicating with them.

Reflecting on the research methodology

Debriefing discussions with fieldworkers and researchers during and after fieldwork indicated what worked well and some learning points.

The children who chose to engage in activities appeared to enjoy them, and the activities generated useful insights. A flexible and empathetic approach to each child was key to achieving this. Fieldworkers responded sensitively to children's needs and preferences; researchers recognised when children were reluctant to join in or wanted to stop, and respected their wishes. They also adapted approaches to reflect children's capabilities and preferences.

Parents valued the opportunity for their children to take part and share their experiences of services. A typical comment included: 'I think that the children should be allowed to. I think it's important that they feel heard, listened to, even at a young age, even at like four they know what they like and what they don't like.' Parents also said that children's observations added insight into how issues (such as low income) directly affected them. Such data is invaluable especially when it concerns topics such as income, not usually discussed with young children.

Some of the activities worked particularly well. For example, using the 'Monkey' soft toy to help children express their ideas and views encouraged children to talk about their homes and their experiences of nursery and visiting the doctor. Subject-specific play equipment and activities were also helpful.

Challenges

The brief was ambitious in its aim to gain information about young children's experiences of three different services. All the more so, within a timescale and budget which allowed for a single visit and one participation research event with children and families in each case, rather than the multiple visits envisaged in the full Mosaic approach. Researchers optimised the time available by planning different activities for each visit, in order to explore the three topic areas.

However, there were some limitations, which point to the benefits of more in-depth fieldwork. For example, pre-visits or multiple visits would have been helpful in building rapport with children, and maximising their engagement in activities and feedback. More visits would have enabled fieldworkers to plan and offer tailored activities suitable for individual children, families, practitioners and within particular settings.

Allowing more time to familiarise children with the activities and equipment may have helped to develop a Mosaic 'picture'. This was especially important for supporting children to use cameras and the tours of the facilities to capture what was important to them. The early years settings involved did not usually conduct participatory feedback activities with children, so the tours were new and unfamiliar to the children. Although the researchers spent some time showing the children how the cameras worked, the children 'played' with the cameras rather than using them to capture what was important to them.

Apart from not carrying out a full range of activities with each child, limited time meant that it was not possible to reflect on the meaning of multiple sources with individual children and parents. The research was successful in involving children in different activities which generated rich feedback. However, if there had been more time, it might have been possible to find out more about individual children, and to explore this with them and their parents to develop the insights further, as advocated by the Mosaic approach.

Finally, there were limitations to the feedback about health and housing services. It was harder to engage some children in health and housing topics compared to early years. Because the children were seen in early years settings, it was easier to give them meaningful prompts.

Conclusions

The study highlights the value of listening work with children, and what can be achieved, depending on time and resources.

It contributes a unique understanding of what is important to children; how they experience services; and what they think these services are for. This is important for understanding how best to engage with children to ensure they have a positive experience of services, which in turn is essential for ensuring that these services are effective in improving children's wellbeing.

The study also highlighted the limitations of listening work with children. It cannot be used to answer questions about the role of specific services in improving wider or more 'global' outcomes in the same way that speaking with parents might. However, children's views can corroborate parents' views, and can also show how issues, which young children are not usually consulted about, (in this case low income) directly affect them.

Whilst some aspects of the Mosaic approach were adapted or not used in this study, the findings also highlight the value of listening work in which only one research visit is possible. Despite the restrictions within this study, we gathered a wealth of information from the children who took part.

References

- Clark A. and Moss, P. (2011). *Listening to young children: the Mosaic approach*, second edition. London: National Children's Bureau.
- Browne, J., Hood, A. and Joyce, R. (2013). *Child and working age poverty in Northern Ireland*. Institute for Fiscal Studies.
- Early Education (2012). *Development matters in the early years foundation stage*. [pdf] Available at: <http://www.foundationyears.org.uk/files/2012/03/Development-Matters-FINAL-PRINT-AMENDED.pdf> [Last accessed 19/6/16].
- Field, F. (2010). *The foundation years: preventing poor children becoming poor adults*. London: HM Government. [pdf] Available at: <http://webarchive.nationalarchives.gov.uk/20110120090128/http://povertyreview.independent.gov.uk/media/20254/poverty-report.pdf> [Last accessed 19/6/16].
- National Children's Bureau (2011). *Putting listening practice at the heart of early years practice: an evaluation of the Young Children's Voice Network*. London: NCB.
- Paley, V. (1997). *The girl with the brown crayon*. London: Harvard University Press.
- Tickell, C. (2011). *The early years: foundations for life, health and learning: an independent report on the early years foundation stage to Her Majesty's Government*. [pdf] Available at: <http://www.educationengland.org.uk/documents/pdfs/2011-tickell-report-eyfs.pdf> [Last accessed 19/6/16].

Preparing transition from child and adolescent mental health services (CAMHS): evaluating a creative participatory approach

Caroline Lee, Cambridge Institute of Public Health, University of Cambridge

Abstract

Researchers sought to jointly develop a preparation programme for transition from child and adolescent mental health services (CAMHS) with young service users and recent leavers from three NHS mental health trusts. Creative research methods were used within two day-long sessions of exercises, discussions and games, and the approach independently evaluated. This article reports these evaluation findings.

In the main, we found that creative and participatory methods offer a safe and relaxed yet fun and stimulating environment, conducive to thinking 'differently' about future mental health support. Young people were able and keen to have a voice in identifying important preparations, mechanisms and processes in transitioning from CAMHS. Challenges to working in this way within a relatively short timescale included institutional and procedural delays affecting recruitment, and resulting in smaller than intended group sizes. The research team endeavoured to minimise the impact of these challenges. Collaboration with NHS partners was a main strength in this project, enabling research findings to contribute directly to a review of policy and practice.

Funding acknowledgement

This research was funded by the National Institute for Health Research (NIHR)'s CLAHRC East of England programme, and approved by the Cambridge University Central Research Ethics Committee. The views expressed are those of the author and not necessarily those of the NHS, the NIHR or the Department of Health.

Introduction

Young people in child and adolescent mental health services (CAMHS) are required, at age 17/18, to either move back to primary care, or onto an adult mental health service. Current service structures offer limited scope to account for a young person's level of maturity, preparedness, functioning, development, wellbeing or readiness for the transition. Consensus is growing that this 'artificial boundary', based on age rather than need, results in a system which is 'weakest where it needs to be strongest' (McGorry et al, 2013). In response, our researchers sought to co-develop, with young service users and recent leavers, a preparation programme for transition from CAMHS, which takes account of these critical facets in a young person's ability to make a 'successful' transition. The key research aims were to identify the most important components to enable a successful transition and to incorporate these into a transition preparation programme (TPP) to assist young people leaving the service.

A participatory method was chosen to encourage young people to be partners in the process, with key data collection via replicable creative workshops led by Tom Mellor, independent creative workshop designer/facilitator and Valerie Dunn, lead researcher, Department of Psychiatry, University of Cambridge.

The challenge was to include activities that were enjoyable, flexible, responsive, capable of exposing diversity of experience and perceptions, and rigorous in quality. This research workshop approach was developed with skilled, artistically-grounded practitioners for a previous project working with vulnerable groups of young people in care (www.youtube.com/watch?v=o17AHhi_fus), and incorporated activities inspired by teaching, drama, community development and group work practices.

Participative research approaches have their roots in community development and recognise the value of researching 'with' rather than 'on' people (Angell et al, 2015). The researchers chose this approach as one suitable to facilitating engagement with, and giving a voice to, vulnerable groups such as young mental health service users (Lushey and Munro, 2014; Calderwood et al, 2015). The team built the design around creative methods, also seen as popular with vulnerable groups, in offering alternative tools to a language-based approach to research (Leitch, 2006). Creative approaches are also proposed as helpful to exploring sensitive material and in exposing complexity and nuance (Eisner, 2008). Methods are said to afford time and space to think, discuss and develop ideas, in comparison with traditional questionnaire-based methods requiring immediate response (Angell et al, 2015).

The 'transitions' research project

Although it had been hoped to also recruit young people not involved in 'user voice' activities as well as from CAMH youth participation networks, the relatively short duration of the project and institutional delays meant this was not possible. Eighteen current and recent CAMHS users, aged 17-21 years, were recruited to the project through the participation coordinators in each of the three NHS Trusts. The role of participation coordinators is to support involvement of young service users in each NHS Trust. In two of the locations, the role was well established and groups had been running for some time, while in one, it was just beginning. Each group was visited by the lead researcher to explain the project; taken through the participant information sheets; and given consent forms to take away and return if they wished to be involved. Participation coordinators were the linchpin of any involvement, and were present at all stages of the project, facilitating meetings, and often supporting young people to attend both practically and emotionally. Young people who took part were given a voucher in recognition of their time, plus travel expenses. They were also offered the opportunity to attend training and undertake a review of 'transitions' literature.

The research was carried out in two stages:

- Stage one of the research was a two-day flexible framework of activities, including games, targeted exercises and discussions, during which participants explored and shared their experiences, ideas and suggestions
- Stage two was a 'production' stage, intended to bring the material together in a draft TPP – the main aim of the project. The main themes and action points were agreed with young people as the work progressed by a verbal group roundup for feedback after each activity, day and workshop, and circulating written summaries after each stage inviting comment as a means of achieving 'ongoing consensus'.

Between the two stages were shorter, two-hour, workshops with clinicians from two NHS Trusts employing some of the creative exercises and involving some of the young participants summarising findings and outlining proposals for change produced during stage one. Table 1 summarises workshop activities, and Box 1 gives an overview of the resulting prototype 'TPP'. For more on workshop activities, including conception, development, implementation and research findings, see forthcoming paper (Dunn and Mellor, 2016).

Table 1: Sample framework of workshop activities

DAY ONE (breaks and lunchtime removed for brevity)	
Introduction and warm-up (10-25 mins)	'Name ball' game
Thinking about transitions and what works in terms of preparation (1 hour approx.)	Thinking about a known transition (e.g. primary to secondary school). Paired work followed by group discussion
What else is going on in young people's lives which might impact on their transition from CAMHS? (1 hour approx.)	'Socks' game with whole group, followed by: 'character creation' in pairs or small groups
How can MH services improve engagement of young people? (1 hour approx.)	'Anti-model': small groups/pairs design a poster/leaflet
CAMHS/AMHS: what are the main differences? Being a young person in an adult service (30-45 mins approx.)	Whole group or discussion in pairs via characters
Round up, consensus, thanks	Group reflect on day, feedback, evaluation questionnaire

Box 1: The young people's transition preparation programme (TPP)

Young people spent two days working together in creative, participatory workshops based on a series of planned exercises lasting a couple of hours each, and designed to get them thinking about specific themes connected to transitioning from child to adult services. Exercises were interspersed with games and breaks. Table 1 illustrates the typical organisation of a workshop day and a few of the games and exercises. Overall, young people thought that mental health service providers under-estimate the anxiety young people feel as they leave CAMHS and, as a result, the amount of preparation many require. Young people agreed, and with researchers drafted, a first outline of a CAMHS TPP, which encompassed:

- Transition information requirements along with recommendations for content, access and methods of delivery
- Staff training suggestions to improve awareness and sensitivity to the specific needs of vulnerable young people transferring to adult mental health services
- Core person-centred aims
- Preparation requirements, activities and tools
- Preparation timings and structure for monitoring and recording progress
- Supports for young people in transition

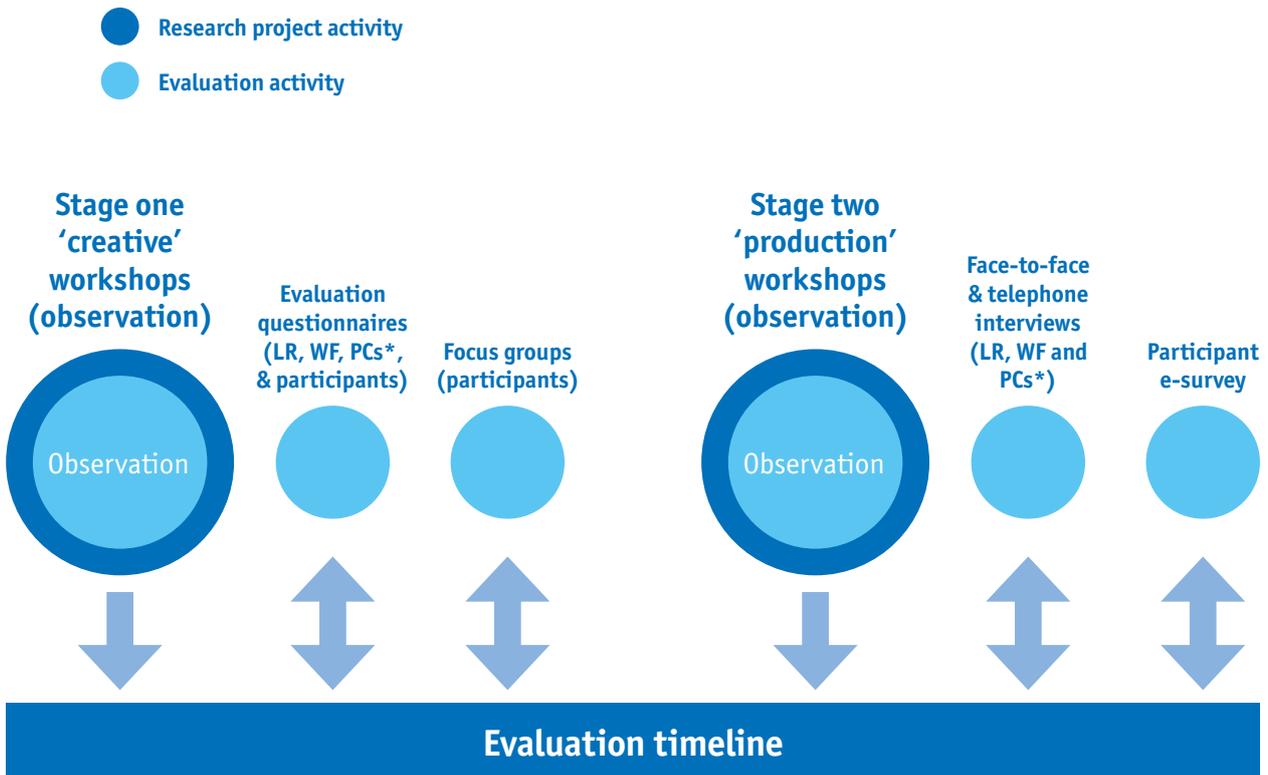
The transitions study opened important dialogues between children and adult mental health services which are set to develop. Contributing to this, 30 clinical staff commented on the perceived barriers to implementing the young people's recommendations in routine NHS mental health settings. Two participating NHS Trusts plan to develop the young people's recommendations.

Thirteen young people took part in stage one workshops: five in location one, and four in locations two and three. Numbers of young people increased overall during stage two. While some participants ceased to be involved, five new participants joined in one location where the NHS Trust participation group was particularly well established. Unfortunately, after the first stage, one NHS Trust (location two) was unable to continue involvement through to stage two. This was due to an overseas opportunity for the young people’s participation group, and both young people and their participation coordinator commented in evaluation questionnaires on how important this project was, and how much they had enjoyed taking part.

Evaluation method

The structure for evaluating the approach is illustrated in Figure 1. It aimed to assess the extent to which the creative, participatory methodology enabled young people to meaningfully collaborate with researchers to produce a prototype TPP. The three lenses of effectiveness, acceptability and feasibility were used as they reflect the value and rationale for selecting creative, participatory methods. That is, are the methods effective at enabling sensitive material to be explored? Do the young people find the approach acceptable and engaging? Is the approach relatively straightforward to implement, and could it be used in similar or different contexts?

Figure 1: Evaluation structure



* Lead researcher (LR), Workshop facilitator (WF), Participation coordinator (PC)

Data collection was based on:

- **Observation** and reflection during workshops, noting on an observation grid (stage one only): engagement with activities, opinions, attitudes and evaluation reflections on the process as well as non-verbal behaviour, atmosphere and adaptations, adjustments made by the lead researcher (LR) or workshop facilitator (WF) to the workshop plan. n=5 (location one), n=4 (location two). Due to scheduling constraints, no independent observation took place in location three
- **Questionnaires** to young people at the end of each day during the stage one workshops (n=13) focusing on preparedness for participation, workshop practicalities and facilities, content/activities and output/impact, triangulated with questionnaires to the LR, WF, and the three participation coordinators (PC). All participants completed evaluation questionnaires at the end of the stage one workshops
- **Focus group** discussion with young people, without researcher or facilitator n=5 (location one), n=4 (location two) to share overall impressions, probe observations noted and elicit opinions and feelings about the activities completed during the two days. Due to scheduling constraints, no focus group took place in location three
- **Semi-structured telephone or face-to-face interviews** after stage two workshops with PCs from locations one and three, the LR and WF reflect fully on the creative, participatory process and explore any themes emerging from questionnaires
- A short **e-survey** issued to participants following stage two (n=8 out of total 18) to gather opinions on the participative process and outcomes, and reflections on involvement in the various activities

The focus was on implementing the creative, participative workshop methodology. Frequency analysis from evaluation questionnaires was used to pinpoint popular and less popular activities, overall value attached to participation by young people and specific elements within the project. Open text from questionnaires contributed to the overall thematic analysis of data from focus groups, feedback sessions and adult interviews. Framework analysis (Gibbs, 2007) was used to code the narrative data into common 'ideas', then to group and organise it according to the three conceptual lenses of effectiveness, acceptability and feasibility.

Findings: effectiveness

Creative methods

In considering effectiveness we were thinking about the extent to which the creative methods enabled young people to share, explore their diverse and shared experiences, and also generate ideas and recommendations. The value of creativity was related strongly to capacity to unlock thoughts and ideas, and the lead researcher reflected on the methodology as one where '(we) are ... creating an environment where creative thinking might take place'. The key to the value of 'creative' methodology, in terms of data generation, was that young people and adults alike emphasised the importance of thinking differently. Participants thought that the workshop exercises provided a conduit for approaching a topic from another direction: 'good to think about things in different ways'; 'helped me think outside the box'.

Hence, ideas may not be so obvious unless first introduced by an exercise, activity, game or product which symbolises or acts as a metaphor for a situation. In reflecting on the popular game: 'Socks!', both young people and participation coordinators agreed that it provided a visual way into discussing the demands of adolescence. One young person was very sceptical of the game at first, and asked what 'the point' was. Afterwards, this participant did see the value and said that it was 'very representative of what is happening (at the time of transition) – lots going on and not in control of anything'. Participation coordinators also saw the importance of this exercise: 'It worked so well because they saw how it represented their lives'.

Whilst arts-based activities could be drawn on, researchers said they were not essential to the method: 'We're using creative in the sense of innovative (or inventive) rather than arty' (LR); '(it's about) different ways to approach finding things out ... bringing on board more arts-based practice, but not necessarily arts' (WF). The young participants also understood this rationale and said how much they valued having a selection of art materials and stationery which gave them, 'flexibility in how to address the activities', even if the activity itself was not overtly 'creative'.

Workshop approach

One of the key factors identified as important by the lead researcher and workshop facilitator to developing a better understanding (of what is going on for the young people) is allocating sufficient time to each activity. The workshop structure afforded young people time and space to reflect, consider, share and explore thoughts and experiences, which is not always there in traditional interview or even focus group situations. The workshop facilitator summarised thus:

'The way of working is a short cut to bringing people out of their shell ... about giving them time to think about. ... and tools to think about it ... (It gives) a different quality to the conversation.'

As well as specific activities within the workshop day, we wanted to understand how the content of the day worked as a programme. Researchers designed a framework with a mix of high-energy (for example throwing and catching socks, moving about the room) and low-energy activities (for example paired discussion, drawing). The participants were positive about this design, stressing the role of games in energising the group: 'high-energy activities make the day better', 'keeps energy levels up, 'really fun, nice and active'.

The majority of participants reported in their questionnaires that the workshops were a good way of doing research. It helped them share thoughts: 'working in a relaxed way enables people to open up'; 'it's a more positive and interactive environment'; 'keeps young people alert and because it's fun we are more interested in giving ideas instead of just sitting and talking'. The participation workers echoed many of these reflections of the young people, reiterating the importance of high and low energy mix: 'getting up and moving around physical space helped with concentration, helped make people feel at ease and helped generate discussion'. Observation data supported this with participants recorded as laughing, giggling and an atmosphere of palpable enjoyment, leading into stimulating, lively discussion.

The mixed evaluation methods exposed how much different groups and individuals enjoyed different activities. For example, while observation notes indicated that participants in all locations were fully engaged in and enjoyed the workshop games, questionnaires and focus groups revealed that some groups and individuals preferred thoughtful discussion whilst others very much preferred to be drawing and writing their thoughts and ideas (for example in a poster or leaflet). Overall indications were that variety and flexibility during the day are vital. The lead researcher and workshop facilitator stressed that having a range of activities to draw on, with a plan B in reserve was important, as was accepting that not all participants enjoy every activity.

Findings: acceptability

Popular and safe

The lens of acceptability focused on the extent to which the young people engaged willingly with the workshop activities, and whether they found them interesting, fun and worthwhile. We found that the fun factor was certainly heavily emphasised by young people and participation coordinators as their experience of the research: '... such an interactive and fun two days'; 'I enjoyed it so much' (participants); '...so much fun in what they were doing, felt like not doing work, even though they were' (participation coordinator). The importance of enjoyment and of young people feeling relaxed was central to the research design and the rationale for choice of method, as the workshop facilitator states: '(it is) difficult to make a one-to-one interview fun, but in games-based workshops, fun is inherent'.

'Feeling safe' was another clear theme to accepting engagement in the research, articulated in both questionnaires and interviews: 'I felt at ease ... and felt that I was able to talk freely' (participant). Creative activities also appeared to offer the possibility to put distance between personal experience – another aspect of methodological 'acceptability', and especially important when question areas are potentially difficult or sensitive. This was again a key influence on the research methods chosen:

'(We) realised the potential (of using arts-inspired practice in research) and see it worked as a good way to talk to more vulnerable groups of people' (WF). Participation coordinators identified that this was indeed what was experienced. Commenting on the 'character creation' exercise, one participation coordinator said it offered a 'safe exploration without becoming too personalised', and another agreed it 'felt like a safe way of talking about difficult things'.

Participatory methods

Calderwood et al (2015) report how 'being listened to' is a key factor in securing participation from young people in research. This was also true of participants in this project, as they themselves articulate:

'Some people find it hard to speak up and get a chance to speak ... it was helpful because everyone could be included and get a chance to say what they wanted'; '... great at keeping us engaged and being patient with us and our individual needs.' 'You guys, like, listen. Didn't feel like research!'

Participation coordinators acknowledged that the young people had a voice in the research: 'seeing the participants expressing their views with increasing confidence, and seeing how their experiences are valued', and reinforced how important being listened to is to young people accepting the process: 'That's why they were engaged – they could see that their voices were being heard (...) young people felt listened to.'

This study aimed to build a meaningful collaboration with young people through an explicit participatory approach. In the stage two workshops, participants decided for themselves which aspects of the TPP they wished to take forward – an aspect highlighted by the lead researcher as a demonstration of young participants influencing research direction, which could also constitute a step up the 'ladder of participation' (Arnstein, 1969). The workshop facilitator confirmed that 'the sessions weren't concrete – they went with the direction that the young people wanted', and participation coordinators corroborated this: 'The approach has been tailored to them and how they want to work'; 'they (participants) could steer the session and that worked really well'.

Findings: feasibility

A strong collaboration with the three NHS Trusts was a key factor in how feasible the research was to undertake in the first place, as well as being a strong determinant of impact.

It is important to reflect on how easy it was for our potential participants to get involved and to remain engaged. Illness/wellness, anxiety, time burden and transport issues can greatly affect research viability. For these reasons, the support role which the participation coordinators provided to young people and the research team proved invaluable, and included advocacy and arranging transport; being aware of each young person's mental wellbeing; and providing emotional support during workshops, when needed.

The collaboration with the NHS Trusts provided insight into the complexity of mental health service pathways. The researchers invited mental health practitioners to a separate workshop so that staff from CAMHS, adult and specialist services could hear feedback from the stage one research, as well as from colleagues in the various mental health services. A participation coordinator summarised the relevance of this: 'A lot of adult and CAMHS knew there were gaps in service, but didn't know how young people were truly feeling at that time'. The young service users also gained understanding from the collaborative aspect, and participants commented on how much they got out of seeing things from the 'other side': 'helped to see it takes a lot of thought to change things', and were surprised by the empathy they felt for therapists: 'made me realise how much services have to think about in order to think about changes. I realised they have a lot to contend with as well'. Although assessing the collaborative aspect of the project was not an explicit evaluation objective, this aspect appeared influential to achieving consensus about what might be achievable in a TPP.

This research was close to practice. Participation coordinators raised the profile of the project and acted as a conduit for the research findings, ultimately enabling the lead researcher to be part of CAMHS transitions review panels in two of the NHS Trusts. As one participation coordinator remarked on the strength of influence of this project's findings: '(it is) quite unusual for research to be so directly fed in ... signs are good compared to other research projects'; 'the research is very much being used – the Trust are now reviewing their transition policy'.

Challenges (and overcoming challenges)

In considering the limitations of the creative, participatory workshop methodology, we acknowledge several challenges, which the project team addressed or minimised the effects of when possible.

Reliability and replicability: the project repeated the creative, participatory methodology in three different NHS Trust locations for specific techniques, (games, exercises, types of questions and prompts). However, the locations and venues were different, start times and arrival times differed, so that each group could have experienced the workshop differently. Further, the make-up of groups was different in each place, reflecting service and pathway differences across locations and differences in NHS patient and public involvement infrastructure. Such differences are inevitable, and rather than aiming for the exact same experience in each location, the techniques used allowed the researchers to respond flexibly to each group within the broad workshop approach. The combination of skilled arts-grounded practitioner with experienced scientific researcher enabled the sort of collaborative team Eisner (2008b) puts forward as a means of enhancing rigour in arts-informed research.

Researcher bias: ethical participatory research suggests that participants should not only have a voice, but also have the opportunity to analyse and interpret research data (Centre for Social Justice and Community Action, 2012). The research design strove to minimise researcher misinterpretation by including young people in analysis, interpretation and reporting, through the ongoing consensus process during and following each stage. It is difficult to conclude with absolute certainty that participants felt they had the capacity or power to offer different interpretations of the data generated. Nevertheless, the evidence supports their agreement with the themes identified for the prototype TPP, as all respondents to the e-survey thought they had contributed to it, and most that they and their ideas had been taken seriously.

Sample bias: the young people who took part were already active in organised patient involvement networks and may not, therefore, be representative of the wider CAMHS population. Ideally, participants would have been recruited over a longer timeframe from more diverse groups of young people but this was not possible within the limitations of a 12-month grant. Specific under-represented groups may have particular preparation needs not uncovered in this research.

Conclusions

The value of the creative, participatory approach lies in bringing together activities and methods which are acceptable and enjoyable to the young people involved in the research, and which are also effective in generating evidence. In this project, the approach demonstrated the immediate value to young people in participating with enjoyment, as well as in generating credible findings with potential to make a difference.

There is clear evidence of the acceptability, and indeed the popularity, of the creative, participatory, workshop method according to indicators of fun, safety, feeling relaxed and listened to, which support the case that it is well suited to working with vulnerable groups. Shared agreement was reached amongst participants and researchers over the content and process for implementing a TPP. Built on strong collaboration, the research project has been able to go that step further. The demonstrable commitment to patient involvement in the three NHS Trusts was facilitated by key individuals from these organisations providing practical support to both the research team and young participants; acting as conduit to awareness within the organisations; and ensuring research is not remote from practice. The production phase was not possible in one of the areas, only because the group became engaged in work overseas.

This evaluation study concludes that a creative, participatory approach can be valuable when researchers require a depth of understanding, nuance and insight, for example about what sort of service change is necessary and why and how it might come about. It can be valuable when the research team and funders are also prepared to allow the research to observe and respond to participants' interests. Stage two of the research explicitly focused on solutions and collaborative ways of working, and meant that there was an in-built intent to bridge the gap of getting research (and policy) into practice (GRIPP), reflecting the rationale behind the CLAHRC programme (www.universitiesuk.ac.uk/aboutus/AssociatedOrganisations/Pages/NIHR-CLAHRC.aspx).

References

- Angell, C., Alexander, J. and Hunt, J.A. (2015). 'Draw, write and tell: a literature review and methodological development on the 'draw and tell' research method.' *Journal of Early Childhood Research* 13(1):17-28.
- Arnstein, S.R. (1969). 'A ladder of citizen participation.' *Journal of the American Institute of Planners* 35(4): 216-224.
- Calderwood, L., Smith, K., Gilbert, E. and Rainsberry, M. (2015). 'Securing participation and getting accurate answers from teenage children in surveys: lessons from the UK Millennium Cohort Study.' *Social Research Practice* 1(1): 27-32.
- Centre for Social Justice and Community Action. (2012). *Community-based participatory research: a guide to ethical principles and practice*. National Co-ordinating Centre for Public Engagement.
- Dunn, V. and Mellor, T. (2016). 'Young people attending and leaving CAMHS in three NHS mental health trusts co-produce a CAMHS Transition Preparation Programme.' (Submitted)
- Eisner, E. (2008). 'Persistent tensions in arts-based research.' In Cahnmann-Taylor, M. and Siegesmund, R. (eds.) *Arts-based research in education: foundations for practice*. Abingdon: Routledge. pp16-27.
- Eisner, E. (2008b). 'Art and Knowledge.' In Knowles, J.G. and Cole, A. L. (eds.) *Handbook of the arts in qualitative research: perspectives, methodologies, examples and issues*. London: SAGE Publishing. pp3-12.
- Gibbs, G.R. (2007). 'Thematic coding and categorizing'. *Analyzing Qualitative Data*. London: SAGE Publishing. pp38-56.
- Leitch, R. (2006). 'Limitations of language: developing arts-based creative narrative in stories of teachers' identities'. *Teachers and teaching: theory and practice*. 12(5): 549-569.
- Lushey, C. and Munro, E. (2014). 'Participatory peer research methodology: an effective method for obtaining young people's perspectives on transitions from care to adulthood?' *Qualitative Social Work* 14(4): 522-537.
- McGorry, P., Bates, T. and Birchwood M. (2013). 'Designing youth mental health services for the 21st century: examples from Australia, Ireland and the UK'. *The British Journal of Psychiatry*.

Addressing quality in social media research: the question of representivity

Josh Keith, Ipsos MORI, Steven Ginnis, Ipsos MORI, Carl Miller, Centre for the Analysis of Social Media, Demos

Abstract

The rise of social media over the past ten years has significant implications for those seeking to conduct research aimed at understanding how society works. Here we discuss the findings from one strand of a wider project tasked with critically examining the commercial possibilities for social media research. This paper explores one of the challenges associated with social media research: the issue of both the representativeness and the reliability of social media analysis. To do this, we review a series of primary research, looking in turn at two different aspects. Firstly, we examine similarities and differences in how attitudes and opinions are voiced through traditional survey research compared to social media (in this case Twitter). Secondly, we examine whether attitudes and opinions expressed are reflective of social media users, in this case all those with Twitter accounts. Finally, we discuss the implications of this for social researchers.

The work discussed here is reported in full in – The road to representivity: a Demos and Ipsos MORI report on sociological research using Twitter: www.ipsos-mori.com/Assets/Docs/Publications/ipsos-mori-demos-road-to-representivity.pdf

Funding acknowledgement

The project that the work in this article is based on is a part of the ‘Wisdom of the Crowd’ project, sponsored by Innovate UK, the UK’s innovation agency, with funding contributions from the TSB, the EPSRC and the ESRC.

We would also like to thank other colleagues at Ipsos MORI, Demos, CASM Consulting LLP, and the University of Sussex who have contributed to this project over the past two years.

More information at: www.ipsos-mori.com/ouexpertise/digitalresearch/sociallistening/wisdomofthecrowd.aspx

Introduction

In this paper, we consider how to add rigour in analysing social media data. In particular, we examine the steps necessary to improve understanding of the profile of a particular social media conversation. In the example presented here, we compare analysis of societal attitudes through Twitter and offline survey datasets.

When measuring public opinion using traditional social research methods, it has always been important to understand how representative any given sample is of the specific population of interest, and therefore, how confident you can be that the research results are representative of the overall population. The work presented in this paper examines the extent to which measurement of attitudes within Twitter datasets are i) representative of specific Twitter conversations; ii) representative of Twitter users; and iii) representative of wider society. The focus is to critically examine how data is collected and analysed to improve understanding and confidence.

Methodological opportunities and challenges

The rise of social media has been rapid. More than six in ten British adults now use social media regularly (Ipsos MORI, 2016). Domo's¹ tracking of the proliferation of data generated every minute suggests that the volume and rate at which we generate data is likely to grow. For example, in 2011, Twitter users sent over 100,000 tweets per minute. This rose to 277,000 in 2013 and to 347,222 in 2015 (Domo, 2015).

This rise of social media presents new opportunities for researchers seeking to understand how society works. Much of the data produced is public and, therefore, readily accessible to researchers in an easily collected digital form. Social media content can be analysed both to improve understanding of the nature of engagement from an anthropological perspective (including networks and language), and as a lens through which to measure attitudes or public opinion.

Such datasets offer tremendous potential. Yet, while relatively straightforward to access and collect, analysing them brings various methodological challenges². Social media datasets are large; reflect attitudes and views in real-time; are passively collected (not requiring the intervention of researchers); and are rich – with links to other online content. This has the potential to allow the researcher to listen to more views, more quickly and in a more natural way than previously possible. Such datasets also allow researchers to delve much deeper into understanding how conversations on social media take place, using techniques grounded in socio-linguistics and network analysis.

While this paper focuses on using Twitter to understand public opinion and attitudes, many of the issues are also relevant to using such data for the other purposes described above.

A central principle of the traditional research conducted in the social sciences, is that the data ought to be representative of the population of interest. However, while we can estimate the proportion of the total population using Twitter, the issue is more complex. Some large national surveys provide information about the profile and demographics of those using various social media platforms, and how frequently. However, this provides only a snapshot of use in time, and does not account for the variation in profile between topics of conversation.

Through our work, we have identified seven ways in which representativeness may pose a challenge for research using Twitter data. Not all of these considerations are as relevant to all forms of social media analysis. However, it is important that all forms of analysis have a basic consideration for each of the following:

1. **Data collection and relevancy** – ensuring a comprehensive body of relevant data is collected
2. **Prolific accounts** – ensuring that individual 'power users' do not dominate analysis through sending a large number of tweets
3. **Bots** – automatically controlled Twitter accounts producing content can distort the conversation
4. **Individuals vs. institutions** – Twitter accounts can represent either individuals or whole institutions, for example news corporations. Accounting for this as part of analysis is a challenge in social media research, for example if trying to capture public vs stakeholder opinion
5. **Location** – placing boundaries on the data, for example tweets generated within the UK, is a particular challenge given the global nature of the overall Twitter conversation about many topics
6. **Socio-demography** – as with survey research, identifying and understanding the socio-demographic profile is important when analysing social media datasets, helping to contextualise any findings
7. **Comparability** – the passive nature of social media data collection compared to active data collection. Social media research does not ask people questions, and therefore, it is measuring different things than survey or qualitative research. This has significant implications for interpreting and framing social media research

¹ www.domo.com

² Increased use of this data for research also brings ethical issues for both data collection and analysis. These need addressing in order to allow the full potential of social media research to be realised, and are subject to a separate strand of the Wisdom of the Crowd project, which presents recommendations for addressing these ethical concerns (Evans et al, 2015).

Methods

Full details of the methods employed as part of the exploration of representativeness on Twitter are in the published report.

Approach to understanding the representativeness of social media data

To understand the problem of representativeness, and evaluate attempts to assess it, we gathered Twitter and survey datasets related to three different topics. The intent was to look at the different ways in which these datasets capture and account for public opinion, and to see how these differences changed when adjustments were made to account for each of the challenges outlined previously. In doing so, the aim was to distinguish between differences produced by the methodology and genuine distinctions in the nature of attitudinal data on Twitter and in response to surveys.

We collected data in each of the following areas:

- **Politics:** did Miliband's and Cameron's survey approval ratings at the height of the general election reflect the support and criticism they received on Twitter? Is Twitter, in general, a harsher or kinder platform, and are there distinctions in how popular leaders were amongst those surveyed, and those who talk about politics on Twitter?
- **Brands:** are certain brands, and certain kinds of brands, more popular in Twitter discussions than those which people mention in surveys?
- **Issues:** do Twitter datasets reflect the issues which survey research suggests are the most important? Do those issues, which people say are the greatest priority when surveyed, also dominate the conversation on Twitter?

Survey data collection

For each theme, we used face-to-face omnibus surveys to measure the views of representative samples. We created scenarios which mimicked the data collected through Twitter as closely as possible. This included incorporating self-completion modules and open-ended questions to minimise interviewer and textual biases. For the purposes of this study, each of the three complex themes were simplified to one main aspect as outlined below.

- **For politics:** respondents were asked a single question about each of the leaders: 'Are you satisfied or dissatisfied with the way David Cameron/Ed Miliband is doing his job as Prime Minister/Leader of the Labour Party?' The first wave was carried out between 8 and 10 February 2015 (asking 1,010 GB adults); the second between 12 and 15 April 2015 (asking 1,000 GB adults).
- **For brands:** a single wave with a representative sample of 961 adults in the UK (aged 15+) was conducted between 5 and 11 June 2015. A single, open question was asked: 'Which brands or companies have you noticed most this week?'
- **For issues:** three survey waves were conducted asking: 'What would you say is the most important issue facing Britain today?' Verbatim responses were coded into categories. Each wave involved a representative sample of adults aged 18+ in Great Britain. The first was conducted between 6 and 12 February 2015 (1,142 respondents); the second between 10 and 20 April 2015 (982 respondents); and the third between 5 and 15 June 2015 (963 respondents).

Twitter data collection

To collect and analyse the Twitter datasets for each theme, we established a research infrastructure as detailed below, attempting to replicate the topics collected through the survey research as closely as possible.

Method52: the project team developed this in partnership with technologists at the University of Sussex and CASM Consulting LLP. It allows non-technical analysts to collect, organise and understand very large datasets, especially those containing text at scales which are too large to manually process. Method52 was mainly used to create and use ‘natural language processing classifiers’. Natural language processing (NLP) combines approaches developed in the fields of computer science, applied mathematics and linguistics. Classifiers are algorithms trained to automatically place tweets in one of a number of pre-defined categories of meaning. Method52 uses NLP technology to allow the researcher to rapidly construct bespoke classifiers to sort defined bodies of tweets into categories (defined by the analyst) based on recognising linguistic differences between different kinds of tweets. Before analysis, the data to be analysed must be selected, and relevant tweets collected, based on a search using one or more keywords as defined by the researcher³. The Twitter API returns a number of fields relating to individual tweets, and to the users responsible for generating those tweets, including a wide range of metadata useful for conducting analysis (for example language, followers, location). Such fields can be sparsely populated, posing issues.

For an example of how the analysis was undertaken, in the politics dataset, two categories of tweets were defined, and all collected, relevant tweets were coded into one of these two datasets:

- ‘Boos’ – tweets identified as being broadly sceptical, or worse
- ‘Cheers’ – tweets identified as being supportive or encouraging

Once classified, it was possible to calculate a ‘net favourability’ score for each leader to compare the survey and Twitter results. This was calculated as the proportion of ‘satisfied’ (survey) or ‘cheers’ (Twitter) minus the proportion of ‘dissatisfied’ (survey) or ‘boos’ (Twitter) for each of David Cameron and Ed Miliband, to give an indication of overall feeling towards the leaders on both media.

Categorisation of data using these classifiers takes place through a phased approach, starting with defining the categories (boos and cheers in the politics example), progressing through a series of training steps whereby the performance of the classifier is reviewed and refined. Once the researcher is satisfied with the performance, the entire dataset is run through the algorithm for classification, and the researcher can progress to post-processing analysis. In this work, classifiers were used to analyse the content of the tweets and the profile of the tweeter (where information was populated), helping with analysis of each of the levels discussed in the following section.

For a more detailed description of this technology, how classifiers are built, and how well they worked in this work, please refer to the full report and appendices.

Once collected and classified using Method52, further analysis of the data was conducted using the Qlikview⁴ data analytics platform to analyse the outcomes of Method52 analysis alongside other pieces of information about individual tweets. This platform acted as a visualisation window into the data processed through Method52, to allow non-technical analysts to identify and explore patterns in the data.

Discussion and conclusions

In this section, we discuss the findings from the primary research. We consider, in turn, the differences observed between two datasets for each topic area; potential contributing factors to the difference; and the implications for conducting social research using social media data.

Addressing issues of representativeness

The research highlighted some important differences between the survey and Twitter data. We highlight some main differences in this section.

³ This identification can either take place before receiving tweets from the Twitter Application Programme Interface (API), if using the ‘search’ API, or after receiving the Twitter dataset, if the ‘stream’ API is being used, which returns a sample of 1% of all tweets sent during a specified period.

⁴ www.qlik.com/products/qlikview

Issues

Analysis showed a difference between the issues spoken about on Twitter and in the survey. For example, ‘crime’ was in the top two on Twitter for all three waves, but was not as prominent in the survey data. In contrast, while ‘immigration’ featured highly in the surveys, it had much lower visibility on Twitter.

There were also some similarities, with ‘education’ and the ‘economy’ featuring fairly prominently in both datasets. Table 1 provides an overview of the similarities and differences across all three waves of the research.

Table 1: Top issues

Issue	February 2015		April 2015		June 2015	
	Twitter rank	Survey rank	Twitter rank	Survey rank	Twitter rank	Survey rank
Crime	1	8	2	8	2	8
Education	2	5	1	4	1	5
Economy	3	3	4	3	4	3
NHS	4	1	3	1	3	2
Defence	5	6	5	7	5	7
Housing	6	7	6	6	6	6
Unemployment	7	4	8	5	8	4
Immigration	8	2	7	2	7	1

Politics

Twitter ‘boos’ and ‘cheers’ for each of David Cameron and Ed Miliband were compared with satisfaction levels when measured through surveys, producing a ‘net favourability’⁵. As shown in Table 2, there was very little relationship between the datasets. Twitter treated both leaders more harshly, with 70% of all tweets sent about the leaders classified as ‘boos’. While Cameron was viewed in a more positive light in the survey data, the Twitter conversation favoured Miliband.

Our analysis showed different trends depending on whether we counted support per tweet, or per person. Both datasets show changes between the two waves, but with negative correlation. While Miliband’s net favourability increased significantly in the surveys, it decreased on Twitter during the same period.

Table 2: Favourability towards party leaders⁶

Leader	February 2015						April 2015					
	Offline			Online			Offline			Online		
	Pos	Neg	Net	Pos	Neg	Net	Pos	Neg	Net	Pos	Neg	Net
Cameron	39	53	-14	16	84	-67	39	53	-14	19	81	-61
Miliband	26	61	-35	26	74	-47	33	52	-19	23	77	-54

⁵ Calculated as proportion ‘satisfied’ minus proportion ‘dissatisfied’ with each leader.

⁶ Percentages may not sum due to computer rounding.

Brands

A comparison of the mentions of top companies and brands on Twitter against unprompted mentions of companies and brands during the survey showed how different the conversation on social media can be to measures of public consciousness offline.

When asked to cite the companies and brands which people have noticed in their everyday conversations offline, no one type of company dominated. Those cited most frequently were banks (9%), supermarkets (9%) and holiday companies or brands (7%), reflecting that most people think about their recent transactional relationships. In contrast, the dynamic of mentions of the top global companies and brands on Twitter was considerably different, and dominated by conversation about technology companies. Mentions captured online reflect that content on social media is proactively generated for a reason; in contrast, most traditional methods for measuring public opinion measure a reactive response to a prompted question. Content generated on Twitter included marketing content, product reviews and news stories.

Table 3: Brand mentions (Twitter)

Brand category	Tweets (n)	Tweets (%)
Tech	163,520	66%
Shops	40,909	16%
Fashion	12,728	5%
Bank	16,642	7%
Supermarket	8,585	3%
Mobile phone network provider	5,175	2%
Food or drink	2174	1%
Overall	249,733	100%

Table 4: Brand mentions (survey)

Brand category	Mentions	Brand category	Mentions
Banks	9%	Mobile phone	1%
Supermarkets	9%	Postal or courier	1%
Holiday companies	7%	Insurance	1%
Shops	5%	Media	1%
Tech	5%	Comparison website	1%
Fashion	4%	Social media	1%
Food or drink	4%	Utilities	1%
Car manufacturers	2%	Other	8%

Understanding and dealing with differences

In the preceding section, we discussed the main ways in which the Twitter and survey data for each of the three themes contrasted. Having established these differences, we then set out to understand how each of seven different layers of potential non-representativeness (in terms of representativeness of the population, the Twitter conversation and people on Twitter) may be contributing to the differences between the two datasets. We suggest that exploration of these themes is important for two reasons: firstly, to understand how each of these sources of non-representativeness (in comparison to survey data) influences any given body of social media data; and secondly, to understand what can currently be done to counteract each issue, if action is required.

While this represents something of a work in progress, here we explain each step and discuss how this worked in practice. This is intended as a useful methodological approach for conducting social research using social media data.

1. Data collection and relevancy

The first challenge is to ensure that data collected really represents what is happening, in this case on Twitter. Tweets are collected based on one or more keywords, as determined by the researcher. This poses two problems. The dataset may include data not relevant to the topic being studied, and relevant data may be missing. Both may introduce systematic bias to the dataset.

Our approach has been to use a broad range of search terms to ensure that all relevant data is collected. This exacerbates the issue of collecting large amounts of irrelevant data, which is more easily dealt with. Training NLP algorithms to recognise the differences in the language between tweets which are relevant and those which are not, allows irrelevant entries to be removed. The accuracy of these algorithms was found to be fairly high across the datasets⁷.

Taking the 'issues' theme, the initial search returned over 14 million tweets. An initial search based on 288 keywords⁸ refined the dataset, and eight of the main issue areas were selected for further analysis. This reduced the dataset to 1.6 million tweets. This body of tweets was then subject to a NLP classifier to sift relevant from irrelevant tweets, reducing the dataset to 129,795 tweets – with the NHS producing the greatest number of irrelevant tweets⁹. This is likely to have been due to the prevalence of words associated with the NHS in other discussions of popular culture, for example, Doctor Who. While this approach is aimed at minimising the amount of relevant data missed, a relevancy classifier should not lead to such a drastic change in a single category, indicating problems with keywords, and that these should be adjusted. In this case, the keywords were not retrospectively adjusted.

2. Prolific accounts

Individual tweets do not represent Twitter users. Across all themes explored in this study, a few 'power users' were responsible for sending a large number of tweets. The most prolific 1% of accounts sent between 14% and 33% of all tweets in the datasets. A reliance on analysis based on number of tweets for a dataset with this level of prolific accounts can skew any analysis if not recognised and accounted for. The dataset most affected was politics. To account for this, users were segmented based on analysis of all the tweets they were responsible for¹⁰. Most users were found to be either very positive or very negative. The influence of prolific users served to boost Ed Miliband's favourability, and decrease that of David Cameron, on Twitter. To take account of this, and base our analysis on individual users rather than tweets, we segmented users based on the overall sentiment of all relevant tweets they generated¹¹. This moved the results closer to the survey satisfaction scores, highlighting the influence on the data of prolific users in this case (see Table 5).

⁷ Accuracy is measured by comparing the performance of computer-categorised content with content manually coded by the researcher.

⁸ Based on the code frame for the survey issues data.

⁹ Over 1.3 million tweets were removed.

¹⁰ 'Very positive' users, 'positive' users, 'undecided' users, 'negative' users and 'very negative' users.

¹¹ If 80-100% of a user's tweets were classified as 'cheers' they were classified as 'very positive'.

Table 5: Influence of power-users on favourability towards Miliband

	February 2015		April 2015	
	Net favourability (tweets)	Net favourability (users)	Net favourability (tweets)	Net favourability (users)
Miliband	-47.6	-54.2	-59.5	-52.5

3. Bots

Not all tweets are posted by human beings, but come from automatically controlled Twitter accounts or 'spambots' intended to appear to represent a large number of individuals active on Twitter. This is one of the more challenging problems to overcome – as analysts find improved ways of identifying bots, their creators work to find ways to make them ever harder to detect. Here we employed a fairly rudimentary approach to filtering out bots, based on identifying accounts with a very low number of followers (less than ten) – although this is just one of several methods that require further consideration. This approach did not identify a large number of bots within any of the three datasets, and therefore, removing these had a negligible impact on the analysis. For example, within the brands dataset, just 1.76% of users had fewer than ten followers, and the tweets sent by this group did not significantly differ from those overall.

4. Individuals vs. institutions

Genuine Twitter accounts can be broadly classified as one of two types: accounts belonging to individuals, or collective accounts belonging to 'institutions'. When seeking to understand attitudes which could differ significantly between these two groups, being able to account for these two types of source in the analysis is important.

Twitter, in particular, offers no easy way to differentiate between the two types of account. Therefore, we trained algorithms to make this distinction in the dataset based on account names and profile fields. The accuracy of this algorithm was tested and found to be accurate 87% of the time.

Once it was possible to account for these two types of account in the analysis, it was found that institutional accounts generated between 10% and 22% of tweets in each of the datasets. However, this had a very small impact on the data, with the exception of politics, where institutional accounts displayed attitudes more similar to survey data than individual accounts.

5. Location

While we expect that Twitter users in the UK may be concentrated more heavily in urban and suburban areas based on research conducted elsewhere (Pew Research, 2015), no robust data for the locational distribution of Twitter users is available.

Only a very small proportion¹² of users include geo-tags within their tweets, allowing location to be easily determined. However, it was possible to use metadata to assign a location to tweets. This performed reasonably well, and was found to be between 80-90% accurate for those tweets it was possible to assign a location to. However, it was not possible to assign location to all tweets¹³.

Examining the regional spread of tweets¹⁴ compared to the distribution of the UK population showed that certain regions are consistently and significantly over-represented, and sometimes demonstrate significantly different opinions. For example, within the politics dataset, London produced 24% of all tweets (10% greater than its share of the population), and Scottish tweeters were more negative towards both leaders (and Miliband in particular).

¹²For example 2.36% of tweets about issues.

¹³For example only 40% of politics tweets could be assigned a location.

¹⁴Based on the 12 NUTS-1 areas.

6. Socio-demography

Socio-demography is an important consideration for ensuring representative survey research – being able to disaggregate by characteristics such as age, gender, socio-economic class and occupation before, or after, data collection. Research shows that Twitter, and social media in general, is not used evenly by all parts of society – with a skew towards men and people in higher socio-economic groups (Ipsos MORI, 2016). It is important to try to identify the demographic profile within a given dataset when examining attitudes which may be driven by these factors.

Our work focused initially on building an algorithm to discern the likely gender split¹⁵ for each of the datasets, while further work will look at other socio-demographic characteristics. For non-institutional accounts, tests showed that this algorithm could operate with 88% accuracy. While men are more likely to use Twitter than women¹⁶, men were over-represented beyond what would be expected, sending 75% of tweets about brands for instance. Furthermore, what people were tweeting about varied by gender – women talked more about education than men – echoing some of the gender distinctions seen in the survey research.

7. Comparability

While not an issue of representativeness, collecting data through social media does not involve asking questions in the same way that survey research does. Therefore, even if steps are taken to ensure the representativeness of a given Twitter dataset, results may differ substantially from survey research on the same topic as they are measuring fundamentally different things. Rather than being subject to any attempted methodological fixes, this requires a clear understanding by the researcher of what they are seeking to achieve through conducting research using social media data, and being transparent about the strengths and weaknesses of this data source for answering any given research question.

Overall conclusions

Researchers wanting to understand society and the way it works cannot ignore the rise of social media usage. Government departments, public services and private sector organisations are all expected to take account of conversations taking place on social media, in the same way it is expected that research is conducted to understand public opinion more widely.

Our work has aimed to understand the methodological challenges associated with conducting such research, and to better equip researchers and the research industry to deal with these. We do not suggest that social media research could replace traditional social research, but our focus is on how best to conduct social media analysis to understand the nature of the comments and opinions voiced, and the nature of the conversation itself.

To this end, we can draw various conclusions from our research. Social media research will never be truly representative, as traditionally understood. However, it is important to seek to understand the different ways in which a social media dataset may differ from a survey dataset; how these differences may vary depending on the topic of investigation; how different sub-sets of the population engage in different conversations on social media; and the impact of each of these on any analysis. In order to move forward from an approach based purely on social media analytics; add rigour to what we do; and move towards a discipline more in line with the traditional principles of social science; it is important to have a common and clear methodological approach to deal with these challenges.

The steps presented in this paper are not intended to be used to make a dataset collected through social media representative of the wider population. The aim is to provide researchers with tools to understand the differences between using social media data and survey research data, and to understand better how this may be used in researching society.

¹⁵Based on individual users being assigned to a gender category.

¹⁶56% of men compared to 44% of women report having visited/used Twitter in the last three months (Ipsos MORI, 2016).

All these steps are work in progress, and require further investigation and refinement as the field continues to develop. Some of the approaches outlined can be useful in other forms of social media analysis, for example, identifying bots, institutional accounts and prolific users would be useful when conducting network analysis of social media conversations.

We also need to challenge assumptions about social media research, and work to inform the design of technological solutions, rather than the other way around. The technology powering social media platforms, and the technology available to analyse the data generated, is constantly evolving. We need to keep up with this, and ensure we are able to lead developments rather than operate within the constraints of wider developments.

References

Evans, H., Ginnis, S., Bartlett, J. (2015). #SocialEthics: a guide to embedding ethics in social media research. London: Ipsos MORI [pdf] Available at <https://www.ipsos-mori.com/Assets/Docs/Publications/im-demos-social-ethics-in-social-media-research-summary.pdf> [Last accessed 19/06/2016].

Ipsos MORI (2016). Technology Tracker Quarterly Release: Q1 2016. [pdf] Available at https://www.ipsos-mori.com/Assets/Docs/Publications/Ipsos_Connect_Tech_Tracker_Q1_2016.pdf [Last accessed 19/06/2016].

James, J. (2015). 'Data Never Sleeps 3.0'. Domosphere. [blog] Available at: <https://www.domo.com/blog/2015/08/data-never-sleeps-3-0/> [Last accessed 19/06/2016].

Miller, C., Ginnis, S., Stobart, R., Krasodomski-Jones, A. and Clemence, M. (2015). The Road to Representivity: A Demos and Ipsos MORI report on sociological research using Twitter. London: Demos. [pdf] Available at <https://www.ipsos-mori.com/Assets/Docs/Publications/ipsos-mori-demos-road-to-representivity.pdf> [Last accessed 19/06/2016].

Pew Research Center (2015). Social Media Update 2014. [pdf] Available at http://www.pewinternet.org/files/2015/01/PI_SocialMediaUpdate20144.pdf [Last accessed 19/06/2016].

Comparison of clustering effects for primary sampling units

Kevin Pickering, Patten Smith and Ben Hornsby, Ipsos MORI

Abstract

The three geographies most used to cluster random probability samples for large scale face-to-face surveys in England are: postcode sectors (PCSs), middle layer super output areas (MSOAs) and Census Area Statistics (CAS) wards. As these three geographies contain, on average, similar numbers of addresses, the clustering effects should be fairly consistent between them. However, changes in the sampling design for some surveys have suggested that this is not the case. We present comparisons of the clustering effects for these three geographies, as well as for a bespoke geography which cuts across them.

Acknowledgements

This study used data from NHS England's GP Patient Survey. We would like to thank Chris Branson, Alex Kong and Will Scott for their contributions to this work.

Background

In order to make the fieldwork for face-to-face surveys manageable (and to contain costs), it is usual to select the samples to be clustered within a random sample of geographical areas¹. This gain in fieldwork efficiency is balanced by a reduction in the statistical efficiency of the sample: it is usually the case that as the average physical size of the geographical area decreases, the fieldwork becomes more efficient, but the statistical efficiency decreases². However, everything else being equal, geographies of a similar size should generate comparable clustering effects.

The three geographies most used previously to cluster random probability samples for large scale face-to-face surveys in England are: postcode sectors (PCSs), middle layer super output areas (MSOAs) and Census Area Statistics (CAS) wards³. It is common practice to merge the smaller postcode sectors, and wards are usually merged with neighbouring ones to ensure a minimum number of addresses. For these analyses we used a minimum of 500 addresses when merging the postcode sectors and wards. The three resulting sets of primary sampling units (PSUs) contain very similar numbers of addresses on average: 3,180 for postcode sectors; 3,400 for MSOAs; and 2,900 for wards. Therefore, we would expect to see similar clustering effects for each.

The formula (Kish, 1965) to estimate the design effect (deff) due to clustering contains two elements: the average number of cases surveyed in the cluster (m) and the rate of homogeneity, usually denoted by ρ :
$$\text{deff (clustering)} \approx 1 + (m-1) \rho$$

The estimate of ρ is dependent on the characteristic being measured and usually ranges between 0 and 1; the higher the value the more similar the units within the cluster are to each other. A value of ρ equal to zero means that the units within the cluster are no more similar to each other than those selected from a simple random sample; a value of 1 means that the characteristic is identical for all units within a cluster.

¹ The samples within selected areas are selected using a systematic (that is 1 in n) sampling approach.

² This loss of statistical efficiency results from units within a cluster being systematically more similar to each other (that is more homogeneous) compared with a simple random sample.

³ For brevity, we refer to these as wards for the rest of this paper.

Kish's formula implies that, if m is kept constant, the design effect has a positive linear relationship with the rate of homogeneity. So, if the same average number of cases is surveyed in each PSU, then the negative impact of the clustering increases as the cases within the PSUs become more similar to each other. As geographical areas shrink, we would expect cases within them to become more similar to each other, and so the rate of homogeneity and hence the design effect to increase. So, for example, we would expect to observe a higher clustering effect if we clustered within a few unit postcodes rather than within whole postcode sectors. Based on this assumption, we would expect to see slightly lower clustering effects for MSOAs, given that they are geographically larger, and slightly higher effects for wards given that they are smaller.

That would be the case if everything was equal. However the Office for National Statistics (ONS) deliberately generated MSOAs to be homogeneous (that is to contain similar households and dwellings). Indeed, every level of the output area (OA) geography was developed to be as homogeneous as possible (Tait, 2012). Wards and postcode sectors, on the other hand, form part of the UK administrative geography – defining areas for elections and used by the Royal Mail for delivering the post. They were not developed to generate homogeneous areas, and however much subtle influences initially and over time might have resulted in a degree of homogeneity, it is highly unlikely that that would have reached the same level as for MSOAs. Therefore, as it is the homogeneity that drives the clustering effect, a priori we would expect to see higher clustering effects for MSOAs than for postcode sectors and wards.

Of these three options, postcode sectors tend to be used as the default geography for defining PSUs for large-scale face-to-face surveys, mainly because the postcode hierarchy makes them easier to work with in practice: knowing that an address is in postcode sector SW12 0 is more informative than that it is in MSOA E02000640 or CAS ward 00AYGS. The fieldwork efficiency and, by implication, costs for these three options would be similar and so would not be a factor in choosing between them.

Wards have been used as the PSUs for three large-scale face-to-face surveys which we know of, although mostly for pragmatic reasons. The first was the Health Survey for England (HSE) 2004⁴ which included a boost sample of participants from black and minority ethnic groups. In order to design the boost sample, area-level ethnic counts were used from the most recent (2001) Census. At the time of sampling, Census 2001 counts by ethnicity were not available for postcode sectors but were available for wards, and so in order to use the most up-to-date information, wards were used to define the PSUs. The second was the Home Office Citizenship Survey (HOCS)⁵ from 2005 to 2010/11. A separate local area study was carried out in 2005, and the Home Office did not want any of the participants in that study to be selected for HOCS 2005. The local area study was carried out in 20 wards, and so, to avoid complexity, it was decided to define wards as the PSUs for HOCS so that those 20 wards could then be simply excluded. Wards were also used as the secondary sampling units (SSUs) for the British Election Panel Study 2009/10⁶, to cluster the sample within the selected parliamentary constituencies, which formed the PSUs.

We are aware of MSOAs being used for two large-scale face-to-face surveys: the British Crime Survey (later the Crime Survey for England and Wales) and Taking Part. For their 2008/9 waves, both the British Crime Survey⁷ and Taking Part⁸ switched from using postcode sectors to adopting a design with a mix of MSOAs and lower layer super output areas (LSOAs). For the 2011/12 wave of the survey, however, Taking Part⁹ reverted to postcode sectors, and for the 2012/13 wave of the Crime Survey for England and Wales¹⁰ a 'bespoke' geography based on combined LSOAs was adopted.

⁴ www.hscic.gov.uk/catalogue/PUB01170/hea-surv-ethn-min-eng-2004-rep-v2.pdf

⁵ http://doc.ukdataservice.ac.uk/doc/5367/mrdoc/pdf/5367_technical_report.pdf

⁶ www.bes2009-10.org/BEStechreportaug2611.pdf

⁷ www.esds.ac.uk/doc/6367/mrdoc/pdf/6367_bcs_2008-09_technical_report_vol1.pdf

⁸ www.gov.uk/government/uploads/system/uploads/attachment_data/file/137862/TPYr4_Technical_Report_November2009.pdf

⁹ www.gov.uk/government/uploads/system/uploads/attachment_data/file/137751/TakingPart-Y7-TechnicalReport.pdf

¹⁰ http://doc.ukdataservice.ac.uk/doc/7422/mrdoc/pdf/7422_csew_2012-13_technical_report.pdf

The technical report for the 2012/13 Crime Survey for England and Wales stated:

'A new bespoke sampling geography was created for the survey. The aim of this was to produce sample clusters that were more heterogeneous than MSOAs...'

This implies that when MSOAs had been used as PSUs for that survey, higher levels of clustering were observed in the survey estimates. Our rationale for carrying out the analyses described in this paper was to obtain reliable evidence that clustering within MSOAs would indeed increase the clustering effect compared with using postcode sectors and wards.

Even given that MSOAs were deliberately derived to be homogeneous, it could be the case that there are properties of postcode sectors and wards which might also have increased homogeneity over time to a similar level to MSOAs. We describe some possible reasons for the sake of example, but there are likely to be multiple factors which influence the level of homogeneity, with different factors influencing different measures.

Postcode sectors were created to be used by the Royal Mail for delivering the post. Even if this had been initially done randomly (which is unlikely), over time the postcode in which a household lives is likely to have had some systemic influence on the characteristics of the householders who move or stay in a particular area. For example, postcode is a factor in how much a property is worth, which is itself correlated with the household income; it is also a criterion in the availability of local schools, which would affect household composition. Wards are part of the administrative geography and in particular, are the level at which local elections are held. It is possible that the political composition of a council affects who lives in the area either because of the tendency to want to live with similar people or through local area policies.

These might seem spurious associations, but the levels of intra-area correlation observed are often small – the clustering effect becomes large because the small intra-PSU similarities get magnified by the number of participants within a PSU. So, a phenomenon that has a seemingly negligible influence on an area could still be reflected in a noticeable increase in the clustering effect.

If it were true that traditional geographies become more homogeneous over time, it implies that if we generated a bespoke geography from scratch, which cut across MSOAs, postcode sectors and wards, we might observe lower clustering effects for it. This was the motivation for generating and testing our bespoke geography, which merged unit postcodes to obtain areas of a similar size to the other three geographies.

We generated the estimates of the clustering effects using the GP Patient Survey (GPPS). The GPPS collects data on patient experiences, attitudes and characteristics from patients who have been registered with a GP practice in England continuously for at least six months and are aged 18 or over (Ipsos MORI, 2015).

The analysis presented in this report was carried out on the combined GPPS datasets from Year 8 Wave 2 (January to March 2014) and Year 9 Wave 1 (July to September 2014). The sampling over the two waves was designed to achieve a minimum level of precision for each GP practice in England, with a combined sample of 2.6 million patients issued to achieve about 880,000 completed questionnaires. This means that the sample covers the whole of England and so, in particular, there are sufficient samples in each geography to obtain accurate estimates of the clustering effects.

Methods

Before adding the geographies to the GPPS dataset, we combined any small postcode sectors which contained fewer than 500 addresses with a coterminous postcode sector(s), to ensure a minimum of 500 addresses in each group of postcode sectors. The original 7,778 postcode sectors were combined to form 7,262 areas. Some merging was also required for wards to ensure a minimum number of 500 addresses, although to a much lesser extent as only 37 wards contained fewer than 500 addresses.

For the purposes of this study, we also generated a bespoke geography by merging coterminous unit postcodes to generate areas, which contained about 3,100 addresses on average, to ensure that they were a similar size to the other three geographies. This merging was done solely based on the distances between the postcodes. If it was the case that the clustering effect increases for traditional geographies over time

(as described above), then we would expect to observe lower clustering effects in this newly defined bespoke geography.

Table 1 shows the average (mean and median), and the minimum and maximum number of addresses for each of the geographies. It also shows the mean geographic area in square kilometres.

Table 1: Profile of the four geographies

	Mean	Median	Minimum	Maximum	Mean area (km ²)
PCSSs	3,182	3,111	500	10,041	17.9
MSOAs	3,403	3,292	1,224	7,011	19.1
Wards	2,900	2,412	500	22,220	16.3
Bespoke	3,131	3,134	2,052	3,146	17.6

A range of questions was selected from the GPPS survey for the analyses, mixed between those explicitly about health and other demographic measures (see Appendices A and B). The analyses were carried out for each geography separately using multi-level linear regression models with individuals nested within PSUs:

$$y_{ij} = a + u_j + e_{ij} \quad ij = \text{individual in PSU } j; j = \text{PSU}$$

The rate of homogeneity ρ was then calculated from the estimated variance terms for u_j and e_{ij} in the model (σ_u^2 and σ_e^2 respectively) as:

$$\rho = \sigma_u^2 / (\sigma_u^2 + \sigma_e^2)$$

Although the outcome measures were coded to be binary, the normal approximation implicit in the regression model gives estimates of ρ consistent with the estimates of design from the standard statistical packages (Stata, SPSS and SAS). It is also the method used by the ONS when it carried out similar analyses to compare design effects for the 1991 Census geographies (Bruce et al, 2001).

Results

The complete set of estimates for the rate of homogeneity (ρ) for the demographic measures is shown in Appendix A, and for the health-related measures, in Appendix B. The means of the estimates of ρ for each geography are shown in Table 2, separately for the demographic and health measures. Because some of the questions included more than one category, we weighted the estimates of ρ so that each question was given equal weight. So each estimate of ρ for a category of age, which had eight categories in total, was given a weight of 0.125, whereas each estimate of ρ for religion, which had three categories, was given a weight of 0.333.

Table 2: Mean of the estimates of ρ

	Demographic	Demographic (excl. ethnicity and religion)	Health measures
PCSSs	0.0483	0.0112	0.0072
MSOAs	0.0526	0.0117	0.0076
Wards	0.0385	0.0108	0.0072
Bespoke	0.0502	0.0110	0.0070

For the demographic measures, the mean estimate of ρ is lowest for wards, implying that they are associated with the lowest clustering effects ($\rho_{\text{mean}} = 0.0385$); the highest estimate is for MSOAs ($\rho_{\text{mean}} = 0.0526$). A large factor in these observed differences is the clustering effects for ethnicity and religion. If we remove ethnicity and religion from the average, then the estimates for the four geographies become relatively closer to each other, although are still lowest for wards ($\rho_{\text{mean}} = 0.0108$) and highest for MSOAs ($\rho_{\text{mean}} = 0.0117$).

For the health measures, the mean estimate of ρ is lowest for the bespoke geography ($\rho_{\text{mean}} = 0.0070$) and is marginally higher for wards and postcode sectors (both $\rho_{\text{mean}} = 0.0072$). The estimate of MSOAs is the highest ($\rho_{\text{mean}} = 0.0076$).

The comparisons of the estimates of the clustering effects, therefore, implies that there is not much difference between wards, postcode sectors and the bespoke geography, although the estimates of ρ are higher on average for MSOAs. To test the consistency of this result across the measures, we ranked the geographies in order for each measure from 1 for the lowest estimate of ρ to 4 for the highest. So, low ranks are 'good' in that they imply a lower clustering effect and higher ranks are 'bad' as they imply a higher clustering effect. Because some of the measures included more than one category, we re-weighted the ranking to give each question equal weight, as when calculating the estimates of the mean of ρ in Table 2. This is why the counts in the following tables are not integers, although each row sums to the total number of questions.

Table 3a shows the weighted counts of ranks for each of the four geographies for all the measures. This table shows that MSOAs had the highest clustering effects for most of the measures, and did not have the lowest rank for any of the measures. This result is consistent when looking at the demographic (Table 3b) and health measures (Table 3c) separately. These analyses also show that the bespoke geography had the lowest clustering effects for the health measures, although for the demographic measures wards, postcode sectors and the bespoke geography were all fairly close to each other.

Table 3a: Relative ranks of ρ for the four geographies: all 23 measures

	Lowest rank 1	2	3	Highest rank 4
PCs	3.8	9.8	9.2	0.1
MSOAs	0.0	1.6	3.6	17.8
Wards	5.3	6.7	6.2	4.8
Bespoke	13.8	4.9	4.0	0.3

Table 3b: Relative ranks of ρ for the four geographies: six demographic measures

	Lowest rank 1	2	3	Highest rank 4
PCs	0.8	3.1	2.1	0.1
MSOAs	0.0	0.3	1.0	4.7
Wards	3.1	1.1	0.9	0.9
Bespoke	2.2	1.5	2.1	0.3

Table 3c: Relative ranks of ρ for the four geographies: 17 health measures

	Lowest rank 1	2	3	Highest rank 4
PCSs	3.1	6.8	7.2	0.0
MSOAs	0.0	1.3	2.6	13.2
Wards	2.3	5.6	5.3	3.8
Bespoke	11.7	3.4	1.9	0.0

Discussion and conclusions

Based on the measures analysed, we would conclude that there is little to choose between using wards and postcode sectors in terms of the levels of clustering. Although the bespoke geography was more likely to have lower estimates of the clustering effect than wards and postcode sectors, the differences tended to be fairly small. Given that the three gave fairly similar estimates, there is no convincing case for choosing any one over the others based solely on the statistical criterion. Therefore, given that they are much easier to work with in practice, we would continue to use postcode sectors to define the PSUs as our preference.

Because the bespoke geography was only associated with marginal reductions in the clustering effect, this suggests that the clustering effects for postcode sectors and wards were not that different from those we would observe from simply cutting up the country into similarly sized geographies with a fairly arbitrary approach. In other words, if there was an increase in homogeneity in pre-defined geographies over time, then that increase is minimal.

In contrast, MSOAs stand out as being different from the others. MSOAs were associated with the highest levels of clustering, although they are geographically larger than the other three geographies and so should, if anything, give lower clustering effects.

Table 4 shows the estimated effective sample sizes (n_{eff}) for the four geographies based on the mean estimates of ρ for the health measures, when the average number of cases surveyed in each cluster ranges from 5 to 15. For example (see right-hand column), for a survey that had a total sample size of 10,000 and achieved about 15 interviews per PSU, the average difference in the effective sample size from clustering with MSOAs rather than postcode sectors would be about 43.

Table 4: Effective sample size (n_{eff}) for a sample size of 10,000

	Average number of interviews in each cluster				
	5	7.5	10	12.5	15
PCSs	9,720	9,552	9,391	9,234	9,083
MSOAs	9,706	9,530	9,361	9,198	9,040
Wards	9,720	9,553	9,391	9,235	9,084
Bespoke	9,729	9,568	9,411	9,259	9,113

So, the evidence suggests that using MSOAs would reduce precision, although not be an amount that was detrimental to estimates in the survey. However, given that MSOAs are larger geographically, on average, than postcode sectors, it does mean that the sample would be both marginally less efficient in both the fieldwork and the statistical precision. In other words, using MSOAs appears to be the lose/lose option when choosing between the competing geographies, although both those losses are likely to be negligible in practice.

These analyses were carried out on measures from a health-related survey and one must always be cautious when generalising results from one survey and subject area. However, given that the same trend was displayed in the general demographic measures, and the a priori reason for expecting more homogeneity within MSOAs, these analyses give fairly convincing evidence that there is no statistical reason for surveys, which currently use postcode sectors or wards as the PSUs, to consider switching to MSOAs.

References

Bruce, S., Lound, C. and Elliot, D. (2001). Designing surveys using variances calculated from census data. National Statistics Methodology Series No. 26.

Ipsos MORI. (2015). GP Patient Survey – technical annex: 2014-15 annual report. [pdf] Available at: http://gp-survey-production.s3.amazonaws.com/archive/2015/July/14-008280-01_Technical%20Annex%202014-2015.pdf. [Last accessed 14/06/2016].

Kish, L. (1965). Survey Sampling. New York: John Wiley & Sons, Inc.

Tait, A. (2012). Changes to output areas and super output areas in England and Wales, 2001 to 2011. Office for National Statistics. [pdf] Available at: <http://www.ons.gov.uk/ons/guide-method/geography/products/census/report--changes-to-output-areas-and-super-output-areas-in-england-and-wales--2001-to-2011.pdf>. [Last accessed 14/06/2016].

APPENDIX A: Estimates of the rate of homogeneity (ρ) for the demographic measures

	Estimate	Geography:			
Measure (question number)	%	PCs	MSOAs	Wards	Bespoke
Age (Q48)					
18 to 24	3.8%	0.0059	0.0062	0.0054	0.0047
25 to 34	8.9%	0.0233	0.0237	0.0198	0.0235
35 to 44	12.2%	0.0125	0.0134	0.0120	0.0129
45 to 54	17.1%	0.0014	0.0018	0.0018	0.0018
55 to 64	19.9%	0.0018	0.0027	0.0024	0.0022
65 to 74	21.1%	0.0108	0.0117	0.0114	0.0111
75 to 84	12.9%	0.0070	0.0077	0.0079	0.0076
85 or over	4.1%	0.0044	0.0045	0.0049	0.0042
Ethnic group (Q49)					
British	82.3%	0.2472	0.2709	0.1904	0.2621
White other	5.6%	0.0476	0.0490	0.0376	0.0478
Other	12.1%	0.2236	0.2519	0.1686	0.2405
Religion (Q58)					
No religion	21.9%	0.0214	0.0223	0.0200	0.0212
Christian	66.4%	0.0525	0.0544	0.0426	0.0520
Other	11.7%	0.1417	0.1576	0.1044	0.1475
Working status (Q50)					
Full-time paid work (30 hours or more each week)	33.0%	0.0139	0.0145	0.0142	0.0134
Part-time paid work (under 30 hours each week)	13.3%	0.0025	0.0025	0.0024	0.0026
Unemployed	4.0%	0.0241	0.0249	0.0183	0.0235
Permanently sick or disabled	4.6%	0.0145	0.0151	0.0130	0.0133
Fully retired from work	35.9%	0.0301	0.0308	0.0294	0.0299
Looking after the home	5.5%	0.0053	0.0055	0.0055	0.0050

Full-time education/doing something else	3.9%	0.0092	0.0091	0.0087	0.0069
Parent or legal guardian (Q53)	19.3%	0.0172	0.0182	0.0165	0.0173
Look after/support family etc for physical or mental health problems in old age (Q56)					
No	79.6%	0.0043	0.0044	0.0046	0.0041
Yes, 1-9 hours a week	11.2%	0.0062	0.0062	0.0067	0.0058
Yes, 10+ hours a week	9.2%	0.0046	0.0051	0.0048	0.0046
Mean (ρ)		0.0483	0.0526	0.0385	0.0502

APPENDIX B: Estimates of the rates of homogeneity (ρ) for the health and health behaviour measures

Measure (question number)	Estimate %	Geography:			
		PCs	MSOAs	Wards	Bespoke
Long-standing health condition (Q30)	54.7%	0.0107	0.0108	0.0106	0.0101
Confidence in managing own health (Q33)					
Very confident	42.0%	0.0044	0.0046	0.0044	0.0042
Fairly confident	50.3%	0.0016	0.0018	0.0015	0.0015
Not very confident/not at all confident	7.7%	0.0098	0.0103	0.0089	0.0098
Medical conditions...					
Angina or long-term heart problem (Q31_2)	7.4%	0.0043	0.0044	0.0044	0.0038
Arthritis or long-term joint problem (Q31_3)	18.7%	0.0111	0.0116	0.0113	0.0110
Asthma or long-term chest problem (Q31_4)	11.3%	0.0034	0.0038	0.0038	0.0035
Cancer in the last 5 years (Q31_6)	4.6%	0.0017	0.0020	0.0021	0.0015
Deafness or severe hearing impairment (Q31_7)	5.6%	0.0039	0.0039	0.0042	0.0038
Diabetes (Q31_8)	10.3%	0.0064	0.0069	0.0064	0.0063
High blood pressure (Q31_10)	26.4%	0.0057	0.0058	0.0058	0.0057
Long-term back problem (Q31_13)	12.0%	0.0043	0.0044	0.0040	0.0039
None of these conditions (Q31_17)	33.3%	0.0110	0.0113	0.0110	0.0104

Health state today...					
Mobility (Q34a)					
I have no problems in walking about	68.1%	0.0174	0.0181	0.0174	0.0163
I have slight problems in walking about	14.7%	0.0021	0.0021	0.0022	0.0020
I have moderate problems in walking about	9.9%	0.0060	0.0064	0.0061	0.0060
I have severe problems in/I am unable to walk(ing) about	7.2%	0.0088	0.0092	0.0084	0.0083
Self-care (Q34b)					
I have no problems washing or dressing myself	87.9%	0.0169	0.0173	0.0158	0.0159
I have slight problems washing or dressing myself	5.6%	0.0043	0.0047	0.0045	0.0046
I have moderate/severe problems/ I am unable to wash or dress myself	6.5%	0.0116	0.0121	0.0108	0.0110
Usual activities (Q34c)					
I have no problems doing my usual activities	68.0%	0.0168	0.0174	0.0167	0.0158
I have slight problems doing my usual activities	16.2%	0.0024	0.0025	0.0027	0.0023
I have moderate problems doing my usual activities	9.6%	0.0064	0.0069	0.0063	0.0063
I have severe problems doing/I am unable to do my usual activities	6.2%	0.0083	0.0088	0.0081	0.0081
Pain/discomfort (Q34d)					
I have no pain or discomfort	44.3%	0.0105	0.0112	0.0107	0.0100
I have slight pain or discomfort	31.2%	0.0024	0.0026	0.0024	0.0025
I have moderate pain or discomfort	16.8%	0.0062	0.0065	0.0062	0.0062
I have severe/extreme pain and discomfort	7.7%	0.0145	0.0149	0.0133	0.0135

Anxiety/depression (Q34e)					
I am not anxious or depressed	67.6%	0.0080	0.0088	0.0081	0.0076
I am slightly anxious or depressed	20.2%	0.0006	0.0007	0.0007	0.0005
I am moderately/severely/extremely anxious or depressed	12.2%	0.0112	0.0121	0.0108	0.0108
Smoking habits (Q55)					
Never smoked	53.1%	0.0170	0.0183	0.0167	0.0163
Former smoker	32.4%	0.0117	0.0124	0.0115	0.0120
Occasional smoker	5.9%	0.0046	0.0053	0.0048	0.0049
Regular smoker	8.6%	0.0148	0.0165	0.0147	0.0141
Mean (ρ)		0.0072	0.0076	0.0072	0.0070

The Social Research Association (SRA)
24-32 Stephenson Way
London NW1 2HX

0207 998 0304
admin@the-sra.org.uk
www.the-sra.org.uk

 [@TheSRAOrg](https://twitter.com/TheSRAOrg)

the-sra.org.uk/journal-social-research-practice